

Partial hyperplane activation for generalized intersection cuts

Aleksandr M. Kazachkov^{*1} Selvaprabu Nadarajah² Egon Balas¹
 François Margot¹

¹*Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA*

²*College of Business Administration, University of Illinois at Chicago,
 Chicago, IL*

Abstract

The generalized intersection cut (GIC) paradigm is a recent framework for generating cutting planes in mixed integer programming with attractive theoretical properties. We investigate this computationally unexplored paradigm and observe that a key hyperplane activation procedure embedded in it is not computationally viable. To overcome this issue, we develop a novel replacement to this procedure called partial hyperplane activation (PHA), introduce a variant of PHA based on a notion of hyperplane tilting, and prove the validity of both algorithms. We propose several implementation strategies and parameter choices for our PHA algorithms and provide supporting theoretical results. We computationally evaluate these ideas in the COIN-OR framework

^{*}akazachk@cmu.edu

on MIPLIB instances. Our findings shed light on the strengths of the PHA approach as well as identify properties related to strong cuts that can be targeted in the future.

1 Introduction

An important aspect of modern integer optimization solvers is the use of cutting planes to strengthen a given formulation [1]. Finding methods for generating stronger general-purpose cutting planes has been an active topic of research in the past decade. As part of this trend, Balas and Margot [7] introduced the *generalized intersection cut* (GIC) paradigm with the motivation of finding stronger cutting planes that also possess favorable numerical properties, such as avoiding numerical inaccuracies that arise in traditional cutting plane approaches [29]. Despite offering some theoretical advantages to other modern cutting planes, GICs have remained unexplored computationally. In this paper, we observe that generating GICs as outlined in [7] is computationally intractable due to the exponential size of the linear program used to generate cuts. We offer a solution by extending the notion of a GIC and devising new algorithms to generate such GICs that scale well with the size of the input data. We prove the validity of our algorithms, provide theoretical results that guide our implementation choices, and perform the first computational investigation with GICs. Our investigation identifies properties related to strong cuts that can be targeted in future approaches within the paradigm.

Let $P := \{x \in \mathbb{R}^n : Ax \geq b, x \geq 0\}$ and $P_I := \{x \in P : x_j \in \mathbb{Z} \text{ for all } j \in I\}$ for a set $I \subseteq \{1, \dots, n\}$, where all data is rational. Let C be a relaxation of P defined by a subset of the inequalities defining P . Balas and Margot [7] produce GICs from a collection of intersection points and rays obtained by intersecting the edges of C with the boundary of a convex set S containing no points from P_I in its interior. The simplest GICs are the *standard intersection cuts* (SICs) [5], which use as C a simple polyhedral cone \bar{C} with apex at a vertex \bar{x} of P . GICs generalize SICs by using a tighter relaxation of P , by *activating* additional hyperplanes

of P that are not included in the description of \bar{C} . However, hyperplane activation poses two computational issues: first, maintaining a description of C becomes challenging, and second, the number of points and rays can quickly grow too large for any practical use.

This paper introduces a new method, called *partial hyperplane activation* (PHA), that addresses the issues with the aforementioned *full* hyperplane activation proposed in [7]. The key insight underlying PHA is that we can forgo a complete description of the polyhedron C by always activating hyperplanes on the initial cone \bar{C} , instead of on an iteratively refined relaxation. We show the PHA approach is not only valid, but also generates a collection of intersection points and rays that grows quadratically in size with the number of hyperplane activations, in contrast to the exponential growth exhibited by the full activation procedure.

Activating a hyperplane on \bar{C} creates new vertices lying at most one edge of P away from \bar{x} . Consequently, intersection points are obtained from edges originating at \bar{x} or these distance 1 vertices. Higher-distance PHA methods, with accompanying higher computational cost, can be easily defined by incorporating hyperplane intersections with edges a larger distance from \bar{x} , eventually recovering the full activation procedure. In other words, while we focus in this paper on the details of a distance 1 PHA procedure, its underlying ideas extend to a hierarchy of PHA procedures delivering progressively "stronger" sets of intersection points.

An important observation related to the effectiveness of PHA is that weak intersection points are created whenever a hyperplane being activated intersects rays of \bar{C} that do not intersect the boundary of S . To mitigate this issue, we introduce a class of *tilted* hyperplanes that can be used to avoid such rays. We also show that tilted hyperplanes offer more control over the number of intersection points generated with essentially no additional overhead, but they require a subtle condition to ensure that these points lead to cuts *valid* for P_I , i.e., to inequalities that do not cut off any points of P_I . We use this idea to design a modified PHA algorithm that creates a point-ray collection that grows linearly with the number of hyperplane activations but quadratically in the number of rays of \bar{C} being cut. The notion

of tilting may also have independent merit outside of the GIC paradigm whenever a \mathcal{V} -polyhedral partial description could be useful, i.e., describing a polyhedron by some of its vertices and edges (see, e.g., [30]).

Implementing PHA or its variant with tilting requires several algorithmic design choices, such as the decision of which particular hyperplanes to activate and which objective functions to use when optimizing over the derived collection of intersection points and rays. Our theoretical contributions provide geometric and structural insights into making some of these choices. Our experiments test the strength of GICs with respect to root node gap closed, a commonly used measure of cut strength, based on three hyperplane activation rules and four types of objective directions.

The first hyperplane activation rule we consider is based on the full activation method, which sequentially activates hyperplanes by pivoting to neighbors of \bar{x} . The second rule targets points that lie deep relative to the SIC. The third rule is motivated by a pursuit of a certain class of intersection points called *final*, relating to facet-defining inequalities for the S_k -closure, which is a relaxation of the split closure obtained by refining P through the addition of one simple split disjunction on a variable x_k .

We then discuss the objective directions evaluated in our implementation, which contribute to a better understanding of the process by which a collection of strong cuts can be generated. The first two sets of objective functions, as with the hyperplane rules, are motivated by finding cuts that improve over the SIC, as well as generalizing the usual objective of maximizing violation with respect to \bar{x} . We then give a result motivating the next set of directions, by showing that the optimal objective value over the S_k -closure can be obtained using only final intersection points and describing how to attain this same value using cuts. The fourth objective function type capitalizes on information that can be used from the typical empirical setup in which multiple split disjunctions are applied in parallel.

Finally, we present the first computational results for any algorithm within the GIC paradigm, implemented in the open source COIN-OR framework [24] and using a set of

benchmark instances from MIPLIB [9]. We find that our PHA approach creates a manageable number of intersection points from which we can generate GICs that effectively generalize SICs: GICs generated from our methodology improve the *integrality gap* closed by one round of SICs on 72.5% of the instances tested, closing on average an additional 5.5% of the gap over SICs for the instances with an improvement. Underlying these results is a detailed analysis of the various hyperplane activation and objective function choices for the linear program used to generate GICs. These experiments build an understanding of structural properties of strong GICs. In particular, we find that the objective functions inspired by final intersection points frequently find the strongest cuts from a given point-ray collection compared to the other objectives tested. However, such final intersection points appear infrequently in the point-ray collections generated by PHA, yielding a concrete direction for future work: to target final intersection points more directly.

Our research on GICs adds to the growing literature that generalizes and extends cutting plane methods based on intersection cuts and Gomory cuts. SICs are equivalent to Gomory cuts [21–23], a relationship surveyed in Conforti et al. [12], and are among the simplest, yet most effective, cuts used in practice. As SICs are generated using information from only one row of the simplex tableau, GICs are a natural generalization by using information from multiple rows. Using such information has been studied extensively in the past decade [4, 13, 15–17, 19, 26]. The focus of these approaches is to obtain stronger cuts by using a better cut-generating set obtained from the optimal solution to the continuous relaxation. GICs instead attempt to strengthen SICs using other rows but the same cut-generating set. The collection of intersection points and rays used to derive GICs can also be seen as defining a linear system for verifying the validity of cutting planes [18].

Broadly, GICs fall into the class of algorithms using the classical theoretical approach of polarity to generate cuts [6]. This includes similarities to local or target cuts [11], which use the polar of a set of points and rays to generate cuts, and to the procedures in [26, 28], which use row generation to individually certify the validity of each generated cut. Though

utilizing some of the same theoretical tools, GICs derive points and rays from a fundamentally different perspective and in a way that immediately guarantees cut validity.

This paper is organized as follows. In Section 2, we revisit full hyperplane activation from [7] and demonstrate the impediment to a practical GIC algorithm inherent in that approach. Section 3 then gives the main contribution of this paper, the PHA scheme. Section 4 shows the idea of tilting and gives an additional algorithm that can be used within the PHA procedure. Implementation choices and supporting theory are discussed in Section 5. Lastly, we give the results of our computational experiments in Section 6. Additional supporting theoretical results are contained in the appendix.

2 Full hyperplane activation

In this section, we show that the full hyperplane activation approach to generate GICs as proposed by Balas and Margot [7] is impractical, which provides the motivation for developing our new methodology in later sections. Given objective vector $c \in \mathbb{Q}^n$, the optimization problem is

$$\min\{c^\top x : x \in P_I\}. \tag{IP}$$

Let (LP) denote the *continuous* or *linear relaxation* of (IP), obtained by removing any integrality restrictions on the variables:

$$\min\{c^\top x : x \in P\}. \tag{LP}$$

Let \bar{x} denote an optimal basic feasible solution to (LP). For simplicity, we assume throughout that P is a full-dimensional pointed polyhedron, but our results extend to the general case with minor modifications.

Let S be a P_I -free convex set, a closed convex set such that its *interior*, denoted $\text{int } S$,

contains no points of P_I , and suppose that $\bar{x} \in \text{int } S$. A commonly used such set is formed from a *simple* split disjunction $(x_k \leq \lfloor \bar{x}_k \rfloor) \vee (x_k \geq \lceil \bar{x}_k \rceil)$ on a variable x_k , $k \in I$:

$$S_k := \{x : \lfloor \bar{x}_k \rfloor \leq x_k \leq \lceil \bar{x}_k \rceil\}.$$

Let $\bar{\mathcal{N}} := \mathcal{N}(\bar{x})$ denote the set of non-basic variables at \bar{x} . Let $\bar{C} := C(\bar{\mathcal{N}})$ be the polyhedral cone with apex at \bar{x} and defined by the n hyperplanes corresponding to the non-basic variables $\bar{\mathcal{N}}$; denote these hyperplanes by $\bar{\mathcal{H}}$. Let $\text{rays}(\bar{C})$ be the rays of \bar{C} .

GICs generalize SICs, as the cone \bar{C} is but one possible relaxation of P . \bar{C} has n extreme rays that can be intersected with $\text{bd } S$ to obtain a set of intersection points $\bar{\mathcal{P}}$ and a set of rays $\bar{\mathcal{R}}$ of \bar{C} that do not intersect $\text{bd } S$, so that

$$\begin{aligned}\bar{\mathcal{R}} &:= \{r \in \text{rays}(\bar{C}) : r \cap \text{bd } S = \emptyset\}, \\ \bar{\mathcal{P}} &:= \{p^r : p^r := r \cap \text{bd } S, r \in \text{rays}(\bar{C}) \setminus \bar{\mathcal{R}}\}.\end{aligned}$$

The intersection points in $\bar{\mathcal{P}}$ and rays in $\bar{\mathcal{R}}$ uniquely define the SIC, $\alpha_0^\top x \geq \beta_0$, obtained from \bar{C} and S [6].

Letting \mathcal{H} denote the set of hyperplanes of P , \bar{C} can be replaced by a tighter relaxation of P , denoted C , obtained by activating a subset of hyperplanes from $\mathcal{H} \setminus \bar{\mathcal{H}}$, i.e., of those that define P but not \bar{C} . Subsequently, edges of C can be intersected with $\text{bd } S$ to obtain a set of intersection points \mathcal{P} , and edges that do not intersect $\text{bd } S$ yield a set of rays \mathcal{R} . In contrast to the situation when using \bar{C} , the number of intersection points and rays obtained from C intersected with $\text{bd } S$ can be strictly greater than n , meaning the point-ray collection defines not just one cut, but a collection of valid cuts.

Theorem 4 in [7] defines valid GICs as these cuts and propose to generate them as any

$(\alpha, \beta) \in \mathbb{R}^n \times \mathbb{R}$ satisfying $\alpha^\top \bar{x} < \beta$ and

$$\begin{aligned} p^\top \alpha &\geq \beta \quad \text{for all } p \in \mathcal{P} \\ r^\top \alpha &\geq 0 \quad \text{for all } r \in \mathcal{R}. \end{aligned} \tag{1}$$

For a fixed β , define $\text{CutRegion}(\beta, \mathcal{P}, \mathcal{R})$ as the vectors α feasible to the above system, i.e., the coefficients for inequalities with lower-bound β that are valid for \mathcal{P} and \mathcal{R} . It suffices to consider $\beta \in \{-1, 0, 1\}$ to obtain all possible cuts.

Unfortunately, the above method requires maintaining the \mathcal{V} -polyhedral description of C . The size of this \mathcal{V} -polyhedral description, as well as the number of rows of (1), as shown in Proposition 1, can grow exponentially large in the number of hyperplanes defining C . Together, these two issues make the full hyperplane activation approach unviable.

Proposition 1. *Let C be formed by adding k_h hyperplanes to the description of \bar{C} , and let $(\mathcal{P}, \mathcal{R})$ denote the point-ray collection obtained from intersecting the edges of C with $\text{bd } S$. The cardinality of \mathcal{P} can grow exponentially large in k_h .*

Proof. Suppose we start with \bar{C} and consider the activation of a halfspace H^+ . Let $\text{rays}(H)$ denote the rays from $\text{rays}(\bar{C})$ that are intersected by H before $\text{bd } S$. For each ray $r \in \text{rays}(H)$ that intersects H , $n - 1 - |\text{rays}(H)|$ new edges are created (not counting the edge back to \bar{x} or the edges between new vertices created by activating H). If $|\text{rays}(H)| \approx n/2$, the size of the new point-ray collection will be $O(n^2)$. The desired result follows inductively. \square

3 PHA

In this section, we propose PHA, an approximate activation method that overcomes the exponential growth in the size of the point-ray collection when using full hyperplane activation. Specifically, our algorithm generates at most $O(k_r^2 k_h)$ intersection points and rays, where k_r and k_h are both parameters; $k_r \leq n$ is the maximum number of initial intersection

points and rays that we remove via hyperplane activations and k_h is the number of activated hyperplanes.

Showing the validity of PHA requires a strict extension of the tools used in [7] to prove the validity of full hyperplane activation. We give this extension in Section 3.1 and employ it in Section 3.2, in which we describe a concrete variant of PHA called PHA₁ and prove its validity. To facilitate reading, Table 1 gives a summary of some of the most frequently used notation that is used in the paper.

Table 1: Summary of frequently used notation.

Notation	Description
\mathcal{H}	Hyperplanes defining P
H_h, H_h^+, H_h^-	$H_h = \{x : a_h^\top x = b_h\}$, $H_h^+ = \{x : a_h^\top x \geq b_h\}$, $H_h^- = \mathbb{R}^n \setminus H_h^+$
$\bar{\mathcal{N}}$	Set of non-basic variables defining \bar{x}
\bar{C}	Simple cone defined by the non-basic variables at \bar{x}
$\text{rays}(\bar{C})$	Rays of \bar{C}
$\bar{\mathcal{H}}$	Hyperplanes defining \bar{C}
$(\mathcal{P}, \mathcal{R})$	Set of points on $\text{bd } S$ and rays not intersecting $\text{bd } S$
$(\bar{\mathcal{P}}, \bar{\mathcal{R}})$	Initial point-ray collection from $\bar{C} \cap \text{bd } S$
$\text{CutRegion}(\bar{\beta}, \mathcal{P}, \mathcal{R})$	Feasible region of (PRLP)
$\mathcal{R}_{\mathcal{P}}$	$\{p - \bar{x} : p \in \mathcal{P}\}$
$\hat{C}(\mathcal{P}, \mathcal{R})$	$\bar{x} + \text{cone}(\mathcal{R}_{\mathcal{P}} \cup \mathcal{R})$
$\mathcal{G}(\mathcal{P}, \mathcal{R})$	$\text{conv}(\mathcal{P}) + \text{cone}(\mathcal{R})$
$\mathcal{H}(r)$	Hyperplanes defining a ray r
$\text{rays}(H_h)$	Rays from $\text{rays}(\bar{C})$ intersected by H_h before $\text{bd } S$

3.1 Proper point-ray collections

We define a point-ray collection as a pair $(\mathcal{P}, \mathcal{R})$ of points and rays obtained from intersecting the polyhedron C with $\text{bd } S$, where C is a relaxation of P . Let \mathcal{K} denote the skeleton of the polyhedron C . Let \mathcal{K}' denote the connected component of $\mathcal{K} \cap \text{int } S$ that includes \bar{x} and \mathcal{K}'' denote the union of the other components of $\mathcal{K} \cap \text{int } S$.

Definition 2. *The point-ray collection $(\mathcal{P}, \mathcal{R})$ is called proper if $\alpha^\top x \geq \beta$ is valid for P_I whenever (α, β) is feasible to (1) and $\alpha^\top v < \beta$ for some $v \in \mathcal{K}'$.*

With this definition, Theorem 3 shows that full hyperplane activation creates a proper point-ray collection, which follows easily from the proof of Theorem 4 in [7].

Theorem 3. *The point-ray collection $(\mathcal{P}, \mathcal{R})$ obtained from intersecting all edges of C with $\text{bd } S$, where C is defined by a subset of the hyperplanes defining P , is proper.*

Given a proper point-ray collection, a GIC is defined as a facet of the convex hull of points and rays, $\mathcal{G}(\mathcal{P}, \mathcal{R}) := \text{conv}(\mathcal{P}) + \text{cone}(\mathcal{R})$, referred to by \mathcal{G} for short, that cuts off a vertex of \mathcal{K}' . Balas and Margot [7] show that these facets can be generated by solving the following linear program formed from the given points and rays and optimized with respect to a given objective direction $w \in \mathbb{R}^n$ and a fixed $\beta \in \{-1, 0, 1\}$:

$$\min\{w^\top \alpha : \alpha \in \text{CutRegion}(\beta, \mathcal{P}, \mathcal{R})\}. \quad (\text{PRLP})$$

In fact, we show something stronger in Lemma 4: the cuts we obtain are not only valid for P_I , but also for $\text{conv}(C \setminus \text{int } S)$. This extends Theorem 3 and will be invoked in proving the results in Section 3.2. For simplicity, we assume that the vertex $v \in \mathcal{K}'$ being cut in Definition 2 is \bar{x} in the rest of the paper unless specified otherwise.

Lemma 4. *Let $(\mathcal{P}, \mathcal{R})$ be the point-ray collection obtained from intersecting all edges of C with $\text{bd } S$. If (α, β) is feasible to (1) and $\alpha^\top \bar{x} < \beta$, then*

$$F := \{x \in C : \alpha^\top x < \beta\} \subseteq \text{int } S.$$

Proof. Assume for the sake of contradiction that $F \setminus \text{int } S$ is non-empty. Then $F \cap \text{bd } S$ is non-empty. We show that this implies an intersection point or ray of C will violate $\alpha^\top x \geq \beta$.

First, note that $\text{cl } F \cap \text{bd } S$ cannot contain any vertex violating the inequality $\alpha^\top x \geq \beta$, as each vertex of $\text{cl } F \cap \text{bd } S$ is an intersection point of C . Hence, as we are also assuming $F \cap \text{bd } S$ is non-empty, the linear program $\min\{\alpha^\top x : x \in \text{cl } F \cap \text{bd } S\}$ must be unbounded. Consider initializing this program at \bar{x} and proceeding by the simplex method using the

steepest edge rule. As the simplex method follows edges of C and it will never increase the objective value, we will either eventually intersect $\text{bd } S$ or end up on a ray of C that is parallel to $\text{bd } S$. This is the desired contradiction: in the former case, this is an intersection point that violates the inequality, and in the latter case, this is a ray of C that does not intersect $\text{bd } S$ and violates the inequality. \square

3.2 Algorithm and validity

We describe the PHA₁ approach for activating a hyperplane partially and prove that it yields a proper point-ray collection. The algorithm proceeds similarly to the full hyperplane activation procedure, in that it activates a hyperplane valid for P , but unlike the procedure from [7], PHA₁ can use any hyperplane valid for P (not necessarily one that defines P), and the activation is performed on the initial relaxation \bar{C} instead of an iteratively refined relaxation that includes previously activated hyperplanes. We show this is not only valid but also computationally advantageous. Part of this advantage comes from that fact that PHA₁ only requires primal simplex pivots in the tableau of the LP relaxation to calculate new intersection points and rays.

The trade-off for the computational advantages of PHA₁ is that it potentially creates a weaker point-ray collection compared to full activation, in the sense that the cuts generated from this collection may be weaker. This is because all vertices created from the partial activations are restricted to being only one pivot away from \bar{x} , i.e., at (edge) distance 1 from \bar{x} along the skeleton of P . For this reason, we refer to Algorithm 1 as PHA₁. A stronger version of PHA₁, such as a valid distance 2 (or higher distance) procedure, is not difficult to create. In other words, one could define a hierarchy of PHA procedures of increasing distance using the principles presented below, eventually leading to the full activation procedure.

We now define some notation used in the algorithm description. For a given $(a_h, b_h) \in \mathbb{R}^n \times \mathbb{R}$, let $H_h := \{x : a_h^\top x = b_h\}$ be a hyperplane such that the corresponding halfspace $H_h^+ := \{x : a_h^\top x \geq b_h\}$ is valid for P . Every element of a proper point-ray collection $(\mathcal{P}, \mathcal{R})$

arises from the intersection of an edge of some relaxation of P with $\text{bd } S$; if the starting vertex of that edge lies on a ray $r^j \in \text{rays}(\bar{C})$, then we say that the point or ray *originates* from r^j . For each $r^j \in \text{rays}(\bar{C})$, let the distance to H_h along the ray r^j starting at \bar{x} be

$$\text{dist}(H_h, r^j) := \begin{cases} (b_h - a_h^\top \bar{x}) / a_h^\top r^j & \text{if } a_h^\top r^j > 0, \\ \infty & \text{otherwise.} \end{cases}$$

We will slightly overload this notation to also refer to the distance along ray r^j to $\text{bd } S$, $\text{dist}(\text{bd } S, r^j)$. This will be the distance along r^j from \bar{x} to the nearest facet of S (or ∞ if no facet of S is intersected). With these definitions, we present Algorithm 1, which takes as an input the polyhedron P , the cut-generating set S , the hyperplane H_h being activated, a set of rays \mathcal{R}_A from $\text{rays}(\bar{C})$ that will be permitted to be cut by H_h (a parameter we will use only later), and any proper point-ray collection to be modified via the activation of H_h . We also assume that along with the point-ray collection, we have kept a history of which edge led to each point and ray to be added to the collection. This is used in the algorithm in step 3 when deciding which points and rays to remove.

Algorithm 1 Distance 1 Partial Hyperplane Activation

Input: Polyhedron P defined by hyperplane set \mathcal{H} ; P_I -free convex set S ; hyperplane H_h valid for P ; set of rays $\mathcal{R}_A \subseteq \text{rays}(\bar{C})$; proper point-ray collection $(\mathcal{P}, \mathcal{R})$ contained in \bar{C} .

- 1: **function** PHA₁($P, S, H_h, \mathcal{R}_A, (\mathcal{P}, \mathcal{R})$)
 - 2: **for all** $r^j \in \mathcal{R}_A$ such that $\text{dist}(H_h, r^j) < \text{dist}(\text{bd } S, r^j)$ **do**
 - 3: Remove all points and rays in $(\mathcal{P}, \mathcal{R})$ that violate H_h^+ and originate from a vertex on r^j but do not lie on or coincide with any ray from $\text{rays}(\bar{C}) \setminus \mathcal{R}_A$.
 - 4: **for all** r^{jhi} originating at vertex $r^j \cap H_h$ **do**
 - 5: **if** $\text{dist}(H, r^{jhi}) < \text{dist}(\text{bd } S, r^{jhi})$ for some $H \in \bar{\mathcal{H}} \setminus \mathcal{H}(r^j)$ **then** Skip this ray.
 - 6: **else if** $r^{jhi} \cap \text{bd } S \neq \emptyset$ **then** Add $p^{jhi} := r^{jhi} \cap \text{bd } S$ to \mathcal{P} .
 - 7: **else** Add r^{jhi} to \mathcal{R} .
 - 8: **return** \mathcal{P} and \mathcal{R} .
-

Theorem 5 states the validity of Algorithm 1 for the special case when all rays of \bar{C} are permitted to be cut. We show how to relax this restriction in Section 4.

Theorem 5. *Algorithm 1 when $\mathcal{R}_A = \text{rays}(\bar{C})$ outputs a proper point-ray collection.*

In the rest of this section, we will build to a proof of Theorem 5 with the help of several useful intermediate results. The validity of PHA_1 follows from proving that the inequalities that can be generated from (PRLP) are valid for a relaxation of P_I related only to \bar{x} and the points and rays in the collection. Unlike the proof of full hyperplane activation, this does not rely on any particular relaxation of P used to obtain the points and rays. Given a point-ray collection $(\mathcal{P}, \mathcal{R})$, let $\mathcal{R}_{\mathcal{P}} := \{p - \bar{x} : p \in \mathcal{P}\}$ and $\widehat{C}(\mathcal{P}, \mathcal{R}) := \bar{x} + \text{cone}(\mathcal{R}_{\mathcal{P}} \cup \mathcal{R})$. That is, $\widehat{C}(\mathcal{P}, \mathcal{R})$ is the polyhedral cone with apex at \bar{x} and rays including \mathcal{R} and a ray from \bar{x} through each intersection point in \mathcal{P} .

Lemma 6. *If $(\mathcal{P}, \mathcal{R})$ is a proper point-ray collection, then $P_I \subseteq \widehat{C}(\mathcal{P}, \mathcal{R})$.*

Proof. Otherwise there exists an inequality valid for \mathcal{G} that cuts both \bar{x} and a point of P_I . This contradicts the assumption that $(\mathcal{P}, \mathcal{R})$ is proper. \square

The next two lemmas state inclusion properties for the sets of cuts obtainable from two different point-ray collections.

Lemma 7. *For $i \in \{1, 2\}$, let \mathcal{P}^i denote a set of points on $\text{bd } S$ and \mathcal{R}^i a set of rays. If for every (α, β) such that α is feasible to $\text{CutRegion}(\beta, \mathcal{P}^2, \mathcal{R}^2)$ and $\alpha^\top \bar{x} < \beta$, α is also feasible to $\text{CutRegion}(\beta, \mathcal{P}^1, \mathcal{R}^1)$, then*

$$\widehat{C}(\mathcal{P}^1, \mathcal{R}^1) \subseteq \widehat{C}(\mathcal{P}^2, \mathcal{R}^2).$$

Proof. Let $\widehat{C}^i := \widehat{C}(\mathcal{P}^i, \mathcal{R}^i)$, $i \in \{1, 2\}$. Assume for the sake of contradiction that there exists an extreme ray r of \widehat{C}^1 that is not in \widehat{C}^2 . If r does not intersect $\text{bd } S$, define $q := r$; otherwise, let $q := r \cap \text{bd } S$, and note that $q \notin \widehat{C}^2$. Then there exists an inequality $\hat{\alpha}^\top x \geq \hat{\beta}$ that separates \bar{x} and q from $\mathcal{G}(\mathcal{P}^2, \mathcal{R}^2)$. This is a contradiction, as $\hat{\alpha}$ is feasible to $\text{CutRegion}(\hat{\beta}, \mathcal{P}^2, \mathcal{R}^2)$ and $\hat{\alpha}^\top \bar{x} < \hat{\beta}$, but it is not feasible to $\text{CutRegion}(\hat{\beta}, \mathcal{P}^1, \mathcal{R}^1)$, since it violates the inequality corresponding to q . \square

Lemma 8. *Let P^1 and P^2 be rational convex polyhedra such that $P^1 \subseteq P^2$, and $\bar{x} \in P^2$. Let \mathcal{P}^i and \mathcal{R}^i denote the intersection points and rays generated from the intersection of P^i with $\text{bd } S$, for $i \in \{1, 2\}$. Then, for a given constant β , a vector α feasible to $\text{CutRegion}(\beta, \mathcal{P}^2, \mathcal{R}^2)$ and cutting \bar{x} will also be feasible to $\text{CutRegion}(\beta, \mathcal{P}^1, \mathcal{R}^1)$.*

Proof. Since $P^1 \subseteq P^2$, we have $P^1 \cap \text{bd } S \subseteq P^2 \cap \text{bd } S \subseteq P^2 \setminus \text{int } S$. By Lemma 4, $P^2 \setminus \text{int } S \subseteq \{x : \alpha^\top x \geq \beta\}$, which completes the proof. \square

Next, we remark on one subtlety of Algorithm 1, in step 2. Namely, whenever the hyperplane being activated intersects a ray of \bar{C} outside of $\text{int } S$, the algorithm skips that ray. In Proposition 9, we show this is valid because such intersections would lead to only *redundant* intersection points, in the sense that these points will never be part of a valid inequality that cuts away \bar{x} . We show the result only for points, as rays are never redundant. Recall that \mathcal{K} denotes the skeleton of the polyhedron C , \mathcal{K}' denotes the connected component of $\mathcal{K} \cap \text{int } S$ that includes \bar{x} , and \mathcal{K}'' denotes the union of the other components of $\mathcal{K} \cap \text{int } S$. Let \mathcal{P}' and \mathcal{P}'' denote the intersection points between $\text{bd } S$ and the edges of the closure of \mathcal{K}' and \mathcal{K}'' , respectively.

Proposition 9. *Any intersection points in $\mathcal{P}'' \setminus \mathcal{P}'$ are redundant.*

Proof. Let α be any feasible solution to $\text{CutRegion}(\beta, \mathcal{P}', \mathcal{R})$ such that $\alpha^\top \bar{x} < \beta$. Suppose for the sake of contradiction that a point $w \in \mathcal{P}'' \setminus \mathcal{P}'$ violates the inequality. Since $w \in \mathcal{K}''$ and $\bar{x} \in \mathcal{K}'$, any path from \bar{x} to w along the skeleton of C must intersect $\text{bd } S$ at some point in \mathcal{P}' . Since C is convex and both \bar{x} and w are cut, there must exist such a path entirely in $\{x : \alpha^\top x < \beta\}$, implying some intersection point in \mathcal{P}' is cut, a contradiction. \square

We now use these results to prove Theorem 5.

Proof of Theorem 5. Let $(\mathcal{P}^*, \mathcal{R}^*)$ be the point-ray collection that Algorithm 1 outputs. Let $\hat{C}^* := \hat{C}(\mathcal{P}^*, \mathcal{R}^*)$. Observe that intersecting the edges of \hat{C}^* with $\text{bd } S$ yields exactly the same point-ray collection $(\mathcal{P}^*, \mathcal{R}^*)$. Then, using Lemma 4 with \hat{C}^* , it follows that all obtainable

cuts will be valid for $\widehat{C}^* \setminus \text{int } S$. To prove the theorem, it suffices to show the inclusion $P_I \subseteq \widehat{C}^*$, which we do next.

From Lemma 6, we have that $P_I \subseteq \widehat{C} := \widehat{C}(\mathcal{P}, \mathcal{R})$, where $(\mathcal{P}, \mathcal{R})$ is the proper point-ray collection given as an input to the algorithm. We need to show that the point-ray collection after activating H_h is proper. Let \mathcal{P}^c and \mathcal{R}^c denote the set of points and rays from $(\mathcal{P}, \mathcal{R})$ that violate H_h^+ .

First we consider intersection points that lie on H_h . Let $P^1 := \widehat{C} \cap H_h$ and $P^2 := \bar{C} \cap H_h$, with \mathcal{P}^i and \mathcal{R}^i the corresponding intersection points and rays of P^i with $\text{bd } S$, for $i \in \{1, 2\}$. Since \mathcal{P} and \mathcal{R} are contained in \bar{C} , $\widehat{C} \subseteq \bar{C}$. As a result, we can apply Lemma 8, which implies that every α (for a β such that $\alpha^\top \bar{x} < \beta$) that is feasible to $\text{CutRegion}(\beta, \mathcal{P}^2, \mathcal{R}^2)$ will also be feasible to $\text{CutRegion}(\beta, \mathcal{P}^1, \mathcal{R}^1)$.

Now we consider the entire halfspace H_h^+ . The set of intersection points of $\widehat{C} \cap H_h^+$ with $\text{bd } S$ is $\overline{\mathcal{P}} := \mathcal{P}^1 \cup \mathcal{P} \setminus \mathcal{P}^c$, and the set of rays is $\overline{\mathcal{R}} := \mathcal{R}^1 \cup \mathcal{R} \setminus \mathcal{R}^c$. In addition, we have that $\mathcal{P}^* = \mathcal{P}^2 \cup \mathcal{P} \setminus \mathcal{P}^c$, and $\mathcal{R}^* = \mathcal{R}^2 \cup \mathcal{R} \setminus \mathcal{R}^c$. Here we are assuming that all of the points in \mathcal{P}^2 belong to the same part of the skeleton of $\bar{C} \cap H_h^+$ that contains \bar{x} . This is without loss of generality as a result of Proposition 9.

Hence, for any β , suppose α is feasible to $\text{CutRegion}(\beta, \mathcal{P}^*, \mathcal{R}^*)$ and satisfies $\alpha^\top \bar{x} < \beta$. Then α will also be feasible to $\text{CutRegion}(\beta, \overline{\mathcal{P}}, \overline{\mathcal{R}})$. By Lemma 7, $\widehat{C}(\overline{\mathcal{P}}, \overline{\mathcal{R}}) \subseteq \widehat{C}^*$. Moreover, since $P_I \subseteq \widehat{C}$ and $P_I \subseteq H_h^+$, we have that $P_I \subseteq \widehat{C} \cap H_h^+$. The intersection points and rays obtained from intersecting $\widehat{C}(\overline{\mathcal{P}}, \overline{\mathcal{R}})$ with $\text{bd } S$ are precisely $\overline{\mathcal{P}}$ and $\overline{\mathcal{R}}$. As a result, by Lemma 6, $P_I \subseteq \widehat{C}(\overline{\mathcal{P}}, \overline{\mathcal{R}})$. It follows that $P_I \subseteq \widehat{C}^*$, as desired. \square

Performing Algorithm 1 k_h times is dramatically more efficient than the $O(n^{k_h})$ complexity of full hyperplane activation. For every activation using Algorithm 1, letting $k_r := |\mathcal{R}_A|$, the hyperplane H_h can cut $O(k_r)$ rays to create $O(k_r^2)$ new intersection points. Therefore, if k_h hyperplanes are activated, the number of intersection points generated by partial activation is $O(k_r^2 k_h)$.

In the presence of rays in the initial point-ray collection, we next describe the role of the

parameter \mathcal{R}_A in PHA_1 , which restricts the set of rays from \bar{C} that can be cut by any activated hyperplane. Edges of C that do not intersect $\text{bd } S$ can be viewed as intersection points infinitely far away, and Balas and Margot [7] show that this view suffices when extending $\text{CutRegion}(\beta, \mathcal{P}, \mathcal{R})$ with only points to a valid system with both points and rays. However, when generating GICs, rays play a fundamentally different role than points. In particular, when activating a new hyperplane cuts some $p \in \bar{\mathcal{P}}$, the new intersection points that are created are either on the SIC or are deeper relative to p (see Theorem 3 in [7]). In contrast, any intersection points created as a result of cutting a ray from $\bar{\mathcal{R}}$ will all lie on the SIC, as shown in Proposition 10.

Proposition 10. *Suppose $r \in \bar{\mathcal{R}}$ is a ray that does not intersect $\text{bd } S$, and H is an activated hyperplane that intersects r at $v := r \cap H$. Denote by Y the set of new rays on H that originate from v . Then the subset of rays Y' in Y that intersect $\text{bd } S$ create intersection points on the hyperplane $\alpha_0^\top x = \beta_0$, where $\alpha_0^\top x \geq \beta_0$ is the SIC from \bar{C} and S . The rays r' in $Y \setminus Y'$ satisfy $\alpha_0^\top r' = 0$.*

Proof. Let $\bar{\mathcal{R}}^c \subseteq \bar{\mathcal{R}}$ denote the set of initial rays cut by H before $\text{bd } S$. When ray r intersects H , it creates $n - 1 - |\bar{\mathcal{R}}^c|$ new rays emanating from v , which we denote by Y . There are $n - 1$ hyperplanes that define r . Any ray $r' \in Y$ lies on $n - 2$ of these hyperplanes, as well as on H . There exists a ray $\hat{r} \in \bar{\mathcal{R}} \setminus \bar{\mathcal{R}}^c$ that also lies on these $n - 2$ facets, and r' can be written as a non-trivial conic combination of r and \hat{r} . Let F denote the two-dimensional face of \bar{C} defined by all conic combinations of r and \hat{r} .

Suppose \hat{r} intersects $\text{bd } S$ and let $\hat{p} := \hat{r} \cap \text{bd } S$. By the remark above, the half-line $L := \hat{p} + \theta r$, $\theta \geq 0$ is contained in $\{x : \alpha_0^\top x = \beta_0\} \cap \text{bd } S$, since $r \in \bar{\mathcal{R}}$ does not intersect $\text{bd } S$. Further, L is also contained in F , which is parallel to r , \hat{r} , and r' . Since r' is a non-trivial conic combination of \hat{r} and r , it intersects $\text{bd } S$. The resulting intersection point is on L , which lies on the SIC.

Now suppose \hat{r} does not intersect $\text{bd } S$. Then $\alpha_0^\top r = \alpha_0^\top \hat{r} = 0$. Since r' is a non-trivial conic combination of r and \hat{r} , it too satisfies $\alpha_0^\top r' = 0$. \square

The consequence of Proposition 10 is that activating hyperplanes on rays of \bar{C} that do not intersect $\text{bd } S$ will create weak points, as these points actually lie on the SIC, not deeper as desired. Thus, the input \mathcal{R}_A to Algorithm 1 will, in our implementation, always be chosen from the set of rays of \bar{C} that intersect $\text{bd } S$, i.e., from $\text{rays}(\bar{C}) \setminus \bar{\mathcal{R}}$. Choosing \mathcal{R}_A arbitrarily may in general lead to invalid cuts; the validity of our choice follows from the results in Section 4. The full details of how we implement PHA₁ to generate cuts are given in Algorithm 2.

Algorithm 2 Generalized Intersections Cuts by PHA₁

Input: Convex rational polyhedron P defined by a set of hyperplanes \mathcal{H} ; set $I \subseteq [n]$ denoting the integer variables, objective coefficient vector c ; classes of objectives \mathcal{O} ; number non-tilted hyperplanes k_h ; fractionality threshold ϵ_{frac} .

- 1: **function** PHA₁CUTGENERATOR($P, I, c, \mathcal{O}, k_h, \epsilon_{\text{frac}}$)
- 2: $\bar{x} \leftarrow$ a solution from $\arg \min \{c^\top x : x \in P\}$.
- 3: $\sigma \leftarrow \{j \in I : \max\{\bar{x}_j - \lfloor \bar{x}_j \rfloor, \lceil \bar{x}_j \rceil - \bar{x}_j\} \geq \epsilon_{\text{frac}}\}$.
- 4: $\mathcal{H}_A \leftarrow \emptyset$; $\mathcal{C} \leftarrow \emptyset$.
- 5: **for** $k \in \sigma$ **do**
- 6: $\mathcal{R}_k^\parallel \leftarrow \{r \in \text{rays}(\bar{C}) : \text{dist}(\text{bd } S_k, r) = \infty\}$; $\mathcal{R}_k \leftarrow \mathcal{R}_k^\parallel$; $\mathcal{P}_k \leftarrow \{r \cap \text{bd } S_k : r \in \text{rays}(\bar{C}) \setminus \mathcal{R}_k\}$.
- 7: **for** $h \in \{1, \dots, k_h\}$ **do**
- 8: **for** $k \in \sigma$ **do**
- 9: Choose a hyperplane H to activate.
- 10: $(\mathcal{P}_k, \mathcal{R}_k) \leftarrow \text{PHA}_1(P, S_k, H, \text{rays}(\bar{C}) \setminus \mathcal{R}_k^\parallel, (\mathcal{P}_k, \mathcal{R}_k))$.
- 11: Add to \mathcal{C} valid cuts by solving (PRLP) with objective types \mathcal{O} , always ensuring \bar{x} is cut.
- 12: **return** \mathcal{C} .

4 PHA₁ with tilted hyperplanes

An outcome of proving the validity of PHA₁ is that the hyperplane given as an input to Algorithm 1 can be an arbitrary valid hyperplane for P , not necessarily one from the hyperplane description of P . One application of this insight is to choose hyperplanes that only cut rays of \bar{C} that intersect $\text{bd } S$, to avoid the weak intersection points that would otherwise be created as shown in Proposition 10. However, it is not practical to search for arbitrary valid hyperplanes that satisfy specific desired properties such as which particular rays of \bar{C}

are or are not cut. Instead, we propose to activate *tilted* versions of hyperplanes defining P , and we show this can be done using information from the original non-tilted hyperplane without a need to explicitly calculate the tilted hyperplane; that is, tilting does not lead to any additional computational burden. A byproduct of the tilting theory is that choosing \mathcal{R}_A as in Algorithm 2 is valid.

In Algorithm 3, we provide an alternative, but not mutually exclusive, approach to Algorithm 2 for using PHA₁. Namely, we provide a way to use tilting to reduce the size of the point-ray collection when the number of rays of \bar{C} that are cut is large, which is desirable when seeking cuts stronger than the SIC. When cutting k_r rays and activating k_h hyperplanes, the algorithm creates $O(k_r k_h^2)$ points and rays, which has a clear advantage to the $O(k_r^2 k_h)$ points and rays that are created from Algorithm 2 when k_r is large. In our implementation in Section 6, we compare the two ideas when used individually and together.

We now define a tilted hyperplane. Let H be a hyperplane defining P that is not tight at \bar{x} . The degenerate case is handled separately, in Appendix B. Any n affinely independent points on H define the hyperplane; to define a tilted version of H , we will only need to change some of these points. We use the intersection of H with the n affinely independent rays of \bar{C} (possibly in their negative directions) to define H .

Suppose $\bar{\mathcal{N}} = \{1, \dots, n\}$ to remove indexing issues. For an arbitrary $d \in \mathbb{R}^n \cup \{\infty\}$, for $j \in \bar{\mathcal{N}}$ and the corresponding $r^j \in \text{rays}(\bar{C})$, if $d_j \notin \{0, \infty\}$, define $v(d_j) := \bar{x} + d_j r^j$, and otherwise, set $v(d_j) := r^j$. Let $\mathcal{V}(d) := \{v(d_j) : j \in \bar{\mathcal{N}}\}$.

Let d^H be a vector of length n with entry j equal to $\min\{\text{dist}(H, r^j), -\text{dist}(H, -r^j)\}$ if r^j or $-r^j$ intersects H , and $d_j^H = 0$ otherwise. Each of the vertices in $\mathcal{V}(d^H)$ is affinely independent from the rest because the rays $\text{rays}(\bar{C})$ are affinely independent. Furthermore, all the rays r^j such that $d_j^H = 0$ are affinely independent, and none of the vertices created lie on these rays. Thus, the the set $\mathcal{V}(d^H)$ of n vertices and rays uniquely define the hyperplane H .

A *tilting* of H is defined via a vector $\delta \in \mathbb{R}^n \cup \{\infty\}$, so that the tilted hyperplane \widetilde{H}

is the unique hyperplane through the vertices and rays given in $\mathcal{V}(d^H + \delta)$. While tilting hyperplanes can be used in many ways in conjunction with PHA₁, arbitrary values of δ would lead to invalid cuts. Indeed, it is clear that δ should be chosen in a way that ensures \widetilde{H} is valid for P . However, in Definition 11, we further restrict the range of values for δ based on our motivation of selectively cutting only certain rays of \bar{C} instead of all of the rays. These tighter limits essentially restrict the tilted hyperplane to intersect each ray $r \in \text{rays}(\bar{C})$ either in the original vertex $H \cap r$, or in the initial intersection point $r \cap \text{bd } S$.

Definition 11. *Let $\text{rays}^c(\bar{C}) \subseteq \text{rays}(\bar{C})$ denote the rays of \bar{C} that have been cut by a set of previously activated hyperplanes. Let H be a hyperplane defining P that is not tight at \bar{x} . Then $\delta \in \mathbb{R}^n \cup \{\infty\}$ defines a targeted tilting of H (with respect to S) if $\delta_j \in \{0, \text{dist}(\text{bd } S, r^j) - d_j^H\}$ for all $j \in \{1, \dots, n\}$ such that $0 < d_j^H < \text{dist}(\text{bd } S, r^j)$ and $r^j \notin \text{rays}^c(\bar{C})$, and $\delta_j = 0$ otherwise.*

The motivation for targeted tilting comes not only from Proposition 10, to avoid cutting the rays in $\bar{\mathcal{R}}$, but also from our desire to find GICs strictly stronger than the SIC from the same P_I -free convex set S . This may require cutting many rays of \bar{C} . However, using Algorithm 2 for this purpose could involve unnecessarily cutting some rays of \bar{C} multiple times. This will happen when we have already cut a certain ray of \bar{C} but a hyperplane being subsequently activated intersects it again; not only might this effort be wasted, but also it may create redundant or weak intersection points. As an alternative, we present the targeted tilting algorithm, Algorithm 3, which is tailored to the task of cutting the initial intersection points and rays more efficiently.

Algorithm 3 focuses on cutting one ray of \bar{C} at a time. The hyperplane chosen for each ray is tilted in step 11 in a way that avoids cutting all other rays, other than the ones that have been previously cut, as required in Definition 11. However, this requirement creates a non-monotonicity when performing targeted tilting, in the sense that cutting more rays of \bar{C} could actually lead to generating weaker cuts. This is due to our observation that cutting a ray multiple times may lead to the addition of weak intersection points to the collection. We

Algorithm 3 Generalized Intersection Cuts by PHA₁ and Targeted Tilting

Input: Polyhedron P defined by a set of hyperplanes \mathcal{H} ; a vertex \bar{x} of P ; indices of fractional integer variables σ ; hyperplane selection criterion \mathcal{SC} ; classes of objectives \mathcal{O} ; set $\mathcal{R}_A \subseteq \text{rays}(\bar{C})$ for the rays of \bar{C} to cut.

```

1: function PHA1CUTGENERATORWITHTARGETEDTILTING( $P, \bar{x}, \sigma, \mathcal{SC}, \mathcal{O}, \mathcal{R}_A$ )
2:   for  $k \in \sigma$  do
3:      $\mathcal{R}_k \leftarrow \{r \in \text{rays}(\bar{C}) : \text{dist}(\text{bd } S_k, r) = \infty\}$ ;  $\mathcal{P}_k \leftarrow \{r \cap \text{bd } S_k : r \in \text{rays}(\bar{C}) \setminus \mathcal{R}_k\}$ .
4:   for  $j \in \{1, \dots, n\}$  do ▷ Assume sorted by decreasing reduced cost
5:     If  $r^j \notin \mathcal{R}_A$ , skip this ray and return to step 4.
6:     for  $k \in \sigma$  do
7:        $\mathcal{H}^j \leftarrow \{H \in \mathcal{H} \setminus \bar{\mathcal{H}} : \text{dist}(H, r^j) < \text{dist}(\text{bd } S_k, r^j)\}$ ,  $\delta \leftarrow \{0\}^n$ .
8:       if  $\mathcal{H}^j \neq \emptyset$  then
9:         Select a hyperplane  $H_j \in \mathcal{H}^j$  according to selection criterion  $\mathcal{SC}$ .
10:        for all  $i \in \{j+1, \dots, n\}$  do
11:           $\delta_i \leftarrow \max\{0, \text{dist}(\text{bd } S_k, r^i) - \text{dist}(H_j, r^i)\}$ .
12:           $(\mathcal{P}_k, \mathcal{R}_k) \leftarrow \text{PHA}_1(P, S_k, H_j, \{r^1, \dots, r^j\} \cap \mathcal{R}_A, (\mathcal{P}_k, \mathcal{R}_k))$ .
13:        if  $\log(j) \in \mathbb{Z}$  or  $j = n$  then
14:          Add to  $\mathcal{C}$  valid cuts by solving (PRLP) with objective types  $\mathcal{O}$ , always ensuring  $\bar{x}$  is cut.

```

counteract this in step 14 of Algorithm 3, by performing cut generation before all hyperplanes have been chosen and activated. To reduce computational effort, this step is only performed $\lceil \log(n) \rceil$ times.

Although targeted tilting may reduce the number of weak intersection points that result from cutting the same rays of \bar{C} repeatedly, it may also miss improving the point-ray collection in certain directions. This creates an inherent trade-off when deciding whether to use PHA₁ as part of Algorithm 2 or Algorithm 3. We discuss this as part of our numerical study in Section 6.

Using $\mathcal{V}(d^H + \delta)$, one can easily compute the closed-form expression for the tilted hyperplane \widetilde{H} , which can then be used in PHA₁ to calculate the new point-ray collection. However, the next theorem shows that when a hyperplane obtained by targeted tilting is being activated, the formula for the tilted hyperplane is not required to be computed. The effect is that throughout PHA₁, all calculations can be performed using H and not \widetilde{H} .

Theorem 12. *Let H be a hyperplane defining P and $\text{rays}^c(\bar{C}) \subseteq \text{rays}(\bar{C})$ the set of rays previously cut, \widetilde{H} be a hyperplane obtained via a valid targeted tilting δ of H , and $\mathcal{R}_A :=$*

$\{r^j \in \text{rays}(\bar{C}) : 0 < d_j^H + \delta_j < \text{dist}(\text{bd } S, r^j)\}$. Given a proper point-ray collection $(\mathcal{P}, \mathcal{R})$, the point-ray collection returned by $\text{PHA}_1(P, S, H, \mathcal{R}_A, (\mathcal{P}, \mathcal{R}))$ is proper.

Proof. We show that $\text{PHA}_1(P, S, H, \mathcal{R}_A, (\mathcal{P}, \mathcal{R}))$ returns an identical point-ray collection to $\text{PHA}_1(P, S, \widetilde{H}, \text{rays}(\bar{C}), (\mathcal{P}, \mathcal{R}))$, which we have already proven is proper in Theorem 5. Let $r^j \in \text{rays}(\bar{C})$. We analyze, case-by-case, the steps of Algorithm 1 when activating \widetilde{H} over all the rays $\text{rays}(\bar{C})$ versus activating H over the rays \mathcal{R}_A .

Case 1: $d_j^H + \delta_j \geq \text{dist}(\text{bd } S, r^j)$. In this case, r^j is not selected in step 2 of Algorithm 1. This is because, when activating H , r^j is not in \mathcal{R}_A , and when activating \widetilde{H} , $\text{dist}(\widetilde{H}, r^j) \geq \text{dist}(\text{bd } S, r^j)$. Note that if $\delta_j > 0$, then we are in this case, so we assume $\delta_j = 0$, i.e., $d_j^H = d_j^{\widetilde{H}}$, for the remainder of the proof.

Case 2: $d_j^H \leq 0$. The ray r^j is again not considered in step 2 of the algorithm, this time because $\text{dist}(H, r^j) = \text{dist}(\widetilde{H}, r^j) = \infty \geq \text{dist}(\text{bd } S, r^j)$.

Case 3: $0 < d_j^H < \text{dist}(\text{bd } S, r^j)$. In this case, the ray is processed in step 2 for both H (it belongs to \mathcal{R}_A) and \widetilde{H} . As we noted, $v^j := r^j \cap \widetilde{H}$ is the same as the intersection $r^j \cap H$. It will be useful in the subsequent discussion to compare the edges that originate at v^j for $\bar{C} \cap \widetilde{H}$ and $\bar{C} \cap H$. Other than the edge from v^j to \bar{x} , each such edge is directed towards a ray $r^k \in \text{rays}(\bar{C}) \setminus \{r^j\}$. Let e_k be the edge of $\bar{C} \cap H$ in the direction of r^k , and let \tilde{e}_k be the edge of $\bar{C} \cap \widetilde{H}$ in that same direction. When $\delta_k = 0$, we have that $e_k = \tilde{e}_k$, in which case the steps of Algorithm 1 relating to this edge will be same regardless of whether \widetilde{H} or H is being activated. Thus, we focus on the case that $\delta_k > 0$, i.e., when $0 < d_k^H < \text{dist}(\text{bd } S, r^k)$ and $d_k^H + \delta_k = \text{dist}(\text{bd } S, r^k)$. Note that this means e_k intersects r^k somewhere in $\text{int } S$.

We now check that the same points and rays are removed in step 3 for both procedures. Consider a point or ray, call it p , that is removed in this step by the activation of \widetilde{H} . Let e be the edge that had resulted in the addition of p to $(\mathcal{P}, \mathcal{R})$. It must be that e originates at a vertex that lies between \bar{x} and v^j , and e does not intersect r^k anywhere in $\text{int } S$. Moreover, e crosses \tilde{e}_k for some edge \tilde{e}_k emanating from v^j . Clearly this remains the case when the corresponding edge e_k coincides with \tilde{e}_k . Else, as we observed, $e_k \cap r^k$ occurs in $\text{int } S$.

Meanwhile, \tilde{e}_k does not intersect r^k in $\text{int } S$. As e_k and \tilde{e}_k both originate at v^j , it follows that in order to cross \tilde{e}_k , e must first cross e_k , which implies that p will also be removed in step 3 when activating H , as desired. This argument is illustrated in Figure 1. Conversely, if a point is not removed by the activation of \tilde{H} , it is easy to see that the point will not violate H either.

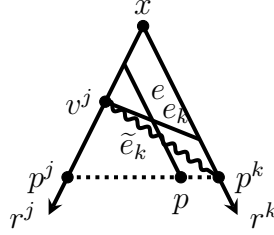


Figure 1: Illustration of case 3 of the proof of Theorem 12. When e_k and \tilde{e}_k do not coincide, it must be the case that e_k intersects r^k in the interior of S and that the point p will be cut by both H and \tilde{H} .

Finally, we show that the same points and rays are added in the for loop starting at step 4. These steps are identical when e_k and \tilde{e}_k coincide, so we again suppose that $e_k \neq \tilde{e}_k$. The activation of \tilde{H} would lead to adding the initial intersection point or ray $r^k \cap \text{bd } S$. However, this point or ray already exists in $(\mathcal{P}, \mathcal{R})$, so it suffices to show that activating H will neither remove this existing point or ray, nor add an extraneous point or ray. The first part is a consequence of the fact that $r^k \notin \mathcal{R}_A$. For the second part, that $e_k \cap r^k$ occurs prior to $\text{bd } S$ means that the condition in step 5 holds, so this edge is skipped and no extraneous points or rays are added to the point-ray collection. \square

One effect of Theorem 12 is that it is valid to simply ignore all intersections of hyperplanes with rays in $\bar{\mathcal{R}}$, as is done in Algorithm 2. To see this, for any hyperplane H valid for P , we can simply set $\delta_j = \infty$ when $d_j^H > 0$ and $\delta_j = 0$ when $d_j^H \leq 0$, resulting in the tilted hyperplane \tilde{H} not intersecting r^j before $\text{bd } S$. The activation performed in step 10 of Algorithm 2 is valid because the rays in $\bar{\mathcal{R}}$ are always ignored, so that the conditions of Theorem 12 apply. Similarly, the theorem applies for step 12 of Algorithm 3.

An arbitrary sequence of valid tilted hyperplanes would not enable us to use this implicit

calculation. Two subtle conditions appearing in Definition 11 are of particular importance for Theorem 12 to hold. First, we require that $\delta_j = 0$ for all rays of \bar{C} that have previously been cut. Second, when δ_j is non-zero, we force it to take the exact value $\text{dist}(\text{bd } S, r^j)$. Appendix C contains an example showing that relaxing either of these conditions may lead to invalid cuts.

5 Implementation choices for PHA₁

Algorithms 2 and 3 both involve steps in which a hyperplane is selected and then, after the points and rays have been collected, a set of objective directions is used for generating GICs from the resulting (PRLP). In this section, we discuss the choices we make for these steps and provide theoretical motivation for the decisions.

5.1 Choosing hyperplanes to activate

5.1.1 Hyperplane activation in Algorithm 3

We first state the three criteria we consider for choosing hyperplanes to activate in step 9 of Algorithm 3, i.e., as the parameter \mathcal{SC} . For an index $k \in I$, we consider optimizing over the S_k -closure, i.e., $\text{conv}(P \setminus \text{int } S_k)$.

(H1) Choose the hyperplane that first intersects some ray $r \in \text{rays}(\bar{C}) \setminus \bar{\mathcal{R}}$ before $\text{bd } S_k$. This corresponds to pivoting to the nearest neighbor from \bar{x} along the ray r , which would be the same hyperplane selected in the procedure from [7].

(H2) Choose the hyperplane that yields a set of points with highest average depth. Here, *depth* is calculated as the Euclidean distance to the SIC (using the same split set). (H2) seeks a set of points that are far from the cut we are trying to improve upon. The idea is that the resulting cuts will then be deeper as well.

(H3) Choose the hyperplane creating the most final intersection points. A *final* intersection point is defined as follows.

Definition 13. Suppose $(\mathcal{P}_k, \mathcal{R}_k)$ is a proper point-ray collection. An intersection point in \mathcal{P}_k or a ray in \mathcal{R}_k is *final* (with respect to S_k) if it belongs to P and $\text{bd } S_k$, meaning it cannot be cut away by any valid hyperplane activations.

We denote the set of final intersection points by \mathcal{P}_k^F and final rays by \mathcal{R}_k^F .

Proposition 14. Suppose $C = P$. Then, the intersection points in \mathcal{P}_k^F define vertices of the S_k -closure and the rays \mathcal{R}_k^F define extreme rays of S_k -closure.

Proof. The S_k -closure is defined as $P_k = \text{conv}(P \setminus \text{int } S_k)$. Therefore, the points $p \in \mathcal{P}^F$ are in the S_k -closure since they belong to $P \cap \text{bd } S_k$. Furthermore, the rays in \mathcal{R}^F are in the S_k -closure since they are extreme rays of P that do not intersect $\text{bd } S_k$.

Consider an arbitrary side of the split disjunction, $S_k^0 := \{x : x_k = \lfloor \bar{x}_k \rfloor\}$. Suppose that a vertex $p = r \cap S_k^0$ in \mathcal{P}_k is not a vertex of the S_k -closure. Then it can be written as a convex combination of other vertices of $P \cap S_k^0$ that are also intersection points created from edges of P . This implies that r , an edge of P , can be written as a conic combination of edges in P that intersect S_k^0 , which is a contradiction. Showing that rays in \mathcal{R}_k are extreme rays of S_k -closure follows similar reasoning. \square

Thus, (H3) targets a set of points that lead to facet-defining inequalities for the set $\text{conv}(P \setminus \text{int } S)$. It turns out to be useful to distinguish between intersection points that are in P and those that are not.

5.1.2 Hyperplane activation in Algorithm 2

Hyperplanes in step 9 of Algorithm 2 are chosen by combining the above options into one rule as follows. For each hyperplane H that is a candidate for activation, we compute the set of intersection points that would result from activating H . We calculate the number of new

final intersection points created, as well as the average depth of the points with respect to the SIC. We choose the best hyperplane H^* based on (H3) and use (H2) as a tie-breaking rule, as several hyperplanes may all create the most final intersection points. We then recompute the number of new final intersection points and average depth of the points for the remaining hyperplanes (after H^* is activated) and repeat the process, until k_h hyperplanes are selected.

5.2 Choosing objective functions

Even if we know that strong cuts exist from a given point-ray collection, it remains to generate these GICs using (PRLP). To do this, we need to appropriately choose objective function coefficient vectors. We consider the following objective directions:

- (R) Ray directions of \bar{C} (i.e., the initial intersection points)
- (V) Vertices that are generated from hyperplane activations
- (T) Intersection points obtained for S_k (the *tight point* heuristic)
- (S) Intersection points from other splits

The first set of objectives, (R), is incentivized by the observation that obtaining GICs that strictly dominate the SIC from the same P_I -free convex set requires cutting many of the initial intersection points. This observation is made concrete in the Theorem 18 in Appendix A; we show that finding a strictly dominating cut requires reducing the dimension of the convex hull of the initial intersection points and rays. Another objective typically used for a cut-generating linear program is to minimize $\bar{x}^\top \alpha$, to obtain the most violated cut with respect to \bar{x} . We generalize this with (V), by optimizing to find the most violated cuts with respect to each of the vertices created on \bar{C} by hyperplane activations.

The third set of directions, (T), is referred to as the tight point heuristic as we are simply trying to find a cut tight on each of the rows of (PRLP) corresponding to points. In other words, we minimize $\alpha^\top p$ for every $p \in \mathcal{P}$ from the current split. Obviously, we will not be able

to cut away p ; the purpose of this objective function is to place a cut as close to the chosen intersection point as possible. The fourth set, (S) , capitalizes on the fact that multiple split sets are considered simultaneously in practice, so that we can share information across the sets to obtain stronger cuts.

The sets of objective directions (T) and (S) switch the perspective typically used for linear programs that generate cuts. Instead of finding a cut with maximal violation, the cuts we obtain from these objectives aim to approximate the convex hull of intersection points and rays from every direction, thereby obtaining as many of the facets of that convex hull as possible and a set of cuts with a wider diversity of angles, which is a desirable computational quality.

Next, we discuss the theoretical motivation for (T) . Consider optimizing over the S_k -closure: $\min\{c^\top x : x \in \text{conv}(P \setminus \text{int } S_k)\}$. We show that intersection points can be used to find an optimal solution to this problem without using (PRLP), and we address how the heuristic (T) aims to obtain the bound implied by this optimal solution.

Theorem 15 shows what bounds can be computed on the optimal value over the S_k -closure using the point-ray collection when $P \subseteq C$, i.e., not all hyperplanes may have been activated. We define $\bar{p}^k \in \arg \min_{p \in \mathcal{P}_k^F} c^\top p$ (when \mathcal{P}_k^F is non-empty) and $\underline{p}^k \in \arg \min_{p \in \mathcal{P}_k} c^\top p$. Let $\bar{z} = c^\top \bar{p}^k$ (defined to be $+\infty$ when \mathcal{P}_k^F is empty) and $\underline{z} = c^\top \underline{p}^k$.

Theorem 15. *Suppose \mathcal{P}_k^F is non-empty. If $\underline{z} < \bar{z}$, then \underline{z} is a lower bound and \bar{z} is an upper bound on the minimum over the S_k -closure. Otherwise, $\bar{z} \leq \underline{z}$ and \bar{z} is the minimum over the S_k -closure.*

Proof. By Proposition 14, \bar{p}^k is in the S_k -closure. Therefore, \bar{z} always provides an upper bound on $c^\top p^*$, where p^* is a minimum over the S_k -closure.

When $\underline{z} < \bar{z}$, $\underline{z} = \min\{c^\top x : x \in C \cap \text{bd } S_k\}$ provides a valid lower bound on $c^\top p^*$, since $P \cap \text{bd } S_k \subseteq C \cap \text{bd } S_k$. On the other hand, when $\bar{z} \leq \underline{z}$, we have $\bar{z} = \min\{c^\top x : x \in C \cap \text{bd } S_k\}$, which implies that \bar{z} provides a lower bound on $c^\top p^*$. Therefore, it must be a minimum over the S_k -closure. \square

Corollary 16 follows immediately for the special case when $C = P$.

Corollary 16. *If $C = P$, then \bar{p}^k is an optimal solution over the S_k -closure.*

Thus, \underline{p}^k is readily available from \mathcal{P}_k and implies a lower bound on the value of the optimal solution over the S_k -closure. The same bound \underline{z} implied by \underline{p}^k can be obtained through GICs, but it may require many cuts generated from (PRLP). We may be able to obtain one inequality tight at \underline{p}^k by using \underline{p}^k itself as the objective coefficient vector, since $\alpha^\top \underline{p}^k \geq \beta$ for all (α, β) feasible to (1), and there will exist a facet-defining inequality $\bar{\alpha}^\top x \geq \bar{\beta}$ for \mathcal{G} satisfying $\bar{\alpha}^\top \underline{p}^k = \bar{\beta}$. Note that for validity of this inequality to be guaranteed, we must additionally verify that the inequality cuts a point $v \in \mathcal{K}'$. However, even if this is satisfied, it is unlikely that the one cut will imply the bound $c^\top x \geq \underline{z}$ on the objective value. In the worst case, n facets of \mathcal{G} that are tight at \underline{p}^k may be required to obtain this bound via cuts. To get these other facets tight at \underline{p}^k , we can use points in \mathcal{P}_k that lie close to \underline{p}^k as the objective vectors. This is precisely what we do in approach (T), though we do not only use \underline{p}^k and points in its vicinity, but also other points from \mathcal{P}_k to encourage diversity of the cut collection.

Finally, we give more details and motivation for the objective directions (S). When using a single split set S_k , no intersection point generated from that split can be cut by any inequality generated through (PRLP) from the point-ray collection for S_k . However, different splits (more generally, different P_I -free convex sets) give rise to different point-ray collections, and the buildup of intersection points and rays can and should be done in parallel for several splits. The first reason for this is computational efficiency: one can intersect an edge with the boundaries of more than one split set and store these intersection points or rays separately. The second reason is that, when using multiple split disjunctions, an intersection point on the boundary of one split set may lie in the interior of another and hence can be cut away by a facet of the point-ray collection from this second split. With this in mind, let σ be the indices of a set of integer variables that are fractional at \bar{x} , i.e., $\sigma \subseteq \{j \in I : \bar{x}_j \notin \mathbb{Z}\}$. For every intersection point p generated from intersecting an edge of C with the boundary

of some split disjunction, we follow the edge to find the *last* split disjunction S_k , $k \in \sigma$, that this edge intersects. If p lies in the interior of S_k , then we will use p as an objective for (PRLP) with feasible region determined by the point-ray collection from S_k .

6 Computational results

This section contains the results from computational experiments with PHA₁ as used in Algorithm 2 from Section 3 and Algorithm 3 from Section 4. The purpose of these experiments is exploratory; that is, we seek conditions and implementation choices that lead to strong GICs. To do this, we evaluate the effect of a variety of parameters, including those discussed in Section 5, and identify structural properties of instances that can be taken advantage of to find stronger cuts. In particular, our results indicate that it can be beneficial to seek a multitude of intersection points that are deep or final as discussed in Section 5.1, and to subsequently target GICs tight on these points through objective functions.

We test Algorithm 2 and Algorithm 3 when used independently, as well as in combination. The purpose of this is to test the strength of targeted tilting relative to non-tilted hyperplanes as in Algorithm 2. When used together, we first perform Algorithm 3 and afterwards perform non-tilted activations. This is equivalent to inserting the steps of Algorithm 3 before step 7 of Algorithm 2.

6.1 Experimental setup

Parameters. Several parameters are fixed throughout the experiments, while others are varied, summarized in Table 2 and elaborated on below.

Table 2: Parameters that are varied and the values considered.

Parameter	Values Considered
Hyperplane scoring function, \mathcal{SC}	{H1, H2, H3}
Objective functions, \mathcal{O}	{ (R) , (V) , (T) , (S) }
Targeted tilting	{On, Off}
# non-tilted hyperplanes, k_h	{0, 1, 2, 3, 4}

The fixed parameters are as follows. The sets used for cut generation are all the simple splits. The time limits are one hour per set of parameters and at most five seconds per objective function for each time (PRLP) is resolved. The instances are not preprocessed, and presolve is turned off for (PRLP), as using presolve adversely affected the quality of the solutions. At most 1,000 GICs are generated per instance (a limit seldom attained). As the cardinality of \mathcal{P} may be large, using all the intersection points as objective directions for (PRLP) may be prohibitively expensive; as a result, we limit the number of points used for objective functions to 1,000 per instance. The overall zero tolerance is set to 10^{-7} , while ϵ_{frac} (used in Algorithm 1) is set to 10^{-3} ; i.e., a point is considered fractional if it is at least 10^{-3} from the nearest integer.

We also perform some standard checks to avoid numerically unstable cuts. Cuts with dynamism higher than 10^6 are rejected, where dynamism is the ratio of the largest and smallest cut coefficients. Additionally, for each candidate cut that is generated, its orthogonality with every existing cut is checked. Given two cuts $\alpha_1^\top x \geq \beta_1$ and $\alpha_2^\top x \geq \beta_2$, their orthogonality is computed as $1 - \frac{\alpha_1 \cdot \alpha_2}{\|\alpha_1\|_2 \|\alpha_2\|_2}$. If the cut already in the pool that is most parallel to the candidate cut has orthogonality less than 0.01 and one of these two cuts is sparser or has a larger Euclidean distance from the LP optimal solution, then only that cut is kept.

While these parameters are fixed, we evaluate the effect of varying which hyperplanes are to be activated and which objective functions are to be used for (PRLP). In Algorithm 2, we vary k_h from 0 to 4, where $k_h = 0$ implies we do not perform any non-tilted activations. Section 6.3 contains computational results relating to the hyperplane activation rules described in Section 5.1. Section 6.4 presents the results for the objective functions discussed in Section 5.2.

Cut generation. We generate cuts in the non-basic space. The vertex \bar{x} of \bar{C} in this space corresponds to the zero vector, and the j th ray of \bar{C} in the non-basic space has a single non-zero j th entry. As a result, the intersection points or rays defining SICs (one pivot from

\bar{x}) and GICs from PHA on \bar{C} (two pivots from \bar{x}) have one and two non-zero coordinates, respectively. Thus, the constraint coefficient matrix of (PRLP) is sparse. This improves the time complexity of implementing Algorithm 1. In addition, it implies that it is sufficient to consider cuts with $\beta = 1$ to obtain all valid inequalities that cut \bar{x} .

Environment. All algorithms are implemented in C++ in the COIN-OR framework [24], using Clp version 1.15 as the linear programming solver. The machine used is a 64-bit PowerEdge R515 with 64GB of memory and twelve AMD Opteron 4176 processors clocked at 2.4GHz. The operating system is Linux Fedora 19 and compiler is g++ version 4.8.3 20140911 (Red Hat 4.8.3-7).

Cut evaluation. The metric we use to evaluate the cuts obtained from a specific set of parameters is *percent gap closed*. The gap is defined as the difference between the optimal values of the integer program and its linear programming relaxation. Denoting the optimal value of the integer program by z_{IP} , of its linear relaxation by z_{LP} , and of the linear relaxation with cuts added by z' , we have

$$\% \text{ gap closed} := 100 \times (z' - z_{LP}) / (z_{IP} - z_{LP}).$$

The baseline we use is the percent gap closed by SICs, which are the simplest GICs. We then add GICs along with the SICs to assess what additional effect GICs have on the percent gap closed in the presence of the SICs.

Instance selection. We test forty instances selected from MIPLIB [2, 9, 10, 25] (all versions) based on the following criteria meant to identify small problems so that many different parameter settings can be tested (over 200 in our experiments): (1) The number of rows and number of columns must be no more than 500 each. (2) The instance has to be integer-

feasible with $z_{LP} < z_{IP}$, and the gap closed by SICs is not 100%. (3) There must be at least one non-final intersection point created from intersecting the rays of \bar{C} with $\text{bd } S$. (4) The instance must not be known to have 0% gap closed from split cuts based on previous experiments [8, 14]. Criterion 3 exists because we do not cut rays of \bar{C} that do not intersect $\text{bd } S$, so if all intersection points are final, no hyperplanes will be activated.

We modify the `stein15`, `stein27`, and `stein45` instances to reduce symmetry, by replacing the objective $\sum_{j=1}^n x_j$ with $\sum_{j=1}^n jx_j$. We also remove the cardinality constraint $\sum_{j=1}^n x_j \geq (n-1)/2$, as this is not present in the initial formulation of these instances [20]. In addition, we exclude instances `go19` and `pp08aCUTS`: the former because its continuous relaxation solves exceptionally slowly, and the latter because it is simply a strengthened version of `pp08a`.

6.2 Point-ray collection statistics

Consider the following statistics averaged across all the instances and all the splits used for each instance, for the best combination of parameters for each instance. The first statistic shows the high prevalence of rays of \bar{C} that do not intersect $\text{bd } S$, which demonstrates the impact of Proposition 10 and our resulting decision to always set the parameter \mathcal{R}_A of Algorithm 1 to a subset of the rays of \bar{C} that intersect $\text{bd } S$. Of the rays of the initial cone \bar{C} , 58% belong to $\bar{\mathcal{R}}$. An additional 5% of the initial rays lead to final intersection points, leaving possibly few rays that can be cut for some instances. For eight instances, only 10% of the rays are able to be cut.

We can also provide an idea of the number of rows in (PRLP) in practice, i.e., the number of points and rays we generate. The number of points ranges from about 21 to 35,000, with an average of 2,350 and median of 217, though there exist instances with split sets leading to as few as 2 and as many as 46,000 points. The number of rays in (PRLP) ranges from 0 to 1,600, with an average of 220 and median of 160. Figure 2 plots the number of generated points across instances (for the best parameter combination for each instance) as a function

of the number of rays of \bar{C} that can be cut; we observe a quadratic relationship, as predicted by our analysis in Section 3.

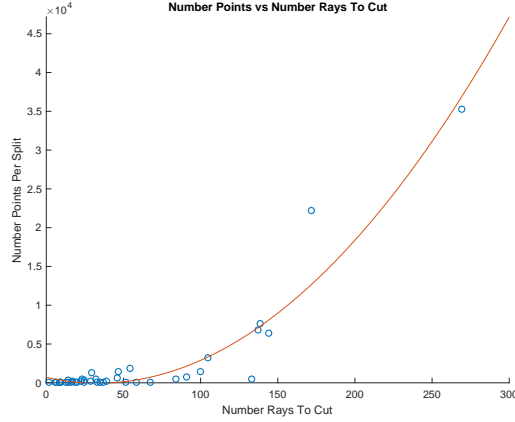


Figure 2: Number generated points versus number of rays that can be cut.

6.3 Effect of hyperplanes activated

In this section, we analyze how the choice of hyperplanes to activate, as described in Section 5.1, affects the percent gap closed by GICs. In addition, we consider whether targeted tilting leads to better cuts. In Section 4, we discussed that, theoretically, neither Algorithm 2 nor targeted tilting is strictly stronger than the other. As a result, computational experiments are necessary to assess the practical strength of each approach.

Table 3 shows the effect of hyperplane activation choices on gap closed for the 29 instances in which GICs close additional gap over SICs. All possible objective functions (discussed in Section 5.2) for (PRLP) are used. The best percent gap closed is shown per instance in: column 2 for SICs; column 3 for GICs across all parameter settings; columns 4 through 7 for Algorithm 2 with $k_h \in \{1, 2, 3, 4\}$; columns 8 through 10 for targeted tilting using each of the hyperplane scoring functions (H1), (H2), (H3); and columns 11 through 14 for when targeted tilting is used in conjunction with Algorithm 2 (for $k_h \in \{0, 1, 2, 3\}$). Note that column T+0 is simply the best result from columns 8 through 10.

The results show that Algorithm 2 closes 98% of the best percent gap closed, and that

Table 3: Best percent gap closed by hyperplane activation choice for instances with any improvement over SICs. Some values differ in the thousandths digit.

Instance	SIC	Best	Only Alg. 2				Only Alg. 3			Alg. 3 with Alg. 2			
			+1	+2	+3	+4	H1	H2	H3	T+0	T+1	T+2	T+3
bell3a	44.74	59.48	59.07	59.07	59.48	59.48	58.94	58.68	58.94	58.94	59.08	59.08	59.08
bell3b	44.57	59.10	59.10	59.10	59.10	59.10	51.36	53.45	51.36	53.45	53.45	53.45	53.45
bell4	23.37	26.52	24.67	24.67	25.26	25.26	23.82	26.35	23.83	26.35	26.35	26.52	26.52
bell5	14.53	85.37	85.37	85.37	85.37	85.37	17.58	19.99	17.58	19.99	19.99	19.99	19.99
blend2	16.00	18.72	16.28	16.28	16.28	16.28	18.72	16.94	18.72	18.72	18.72	18.72	18.72
bm23	5.92	10.98	8.00	10.55	10.98	10.98	8.88	9.94	8.94	9.94	9.94	9.94	9.94
egout	51.57	51.60	51.57	51.57	51.57	51.57	51.59	51.60	51.59	51.60	51.60	51.60	51.60
k16x240	7.56	7.71	7.71	7.71	7.71	7.71	7.56	7.65	7.56	7.65	7.65	7.65	7.65
lseu	4.57	4.65	4.57	4.65	4.65	4.65	4.65	4.57	4.65	4.65	4.65	4.65	4.65
mas74	3.30	4.31	4.30	4.31	4.31	4.31	4.11	3.81	3.98	4.11	4.11	4.11	4.11
mas76	2.37	2.46	2.38	2.46	2.46	2.46	2.37	2.37	2.37	2.37	2.37	2.37	2.37
mas284	0.38	0.41	0.38	0.38	0.38	0.38	0.41	0.41	0.41	0.41	0.41	0.41	0.41
misc05	3.60	4.48	3.60	3.60	3.60	3.60	3.60	4.48	3.60	4.48	4.48	4.48	4.48
mod008	1.31	1.38	1.31	1.38	1.38	1.38	1.31	1.31	1.31	1.31	1.31	1.31	1.31
mod013	4.41	7.37	4.41	4.72	4.72	7.28	4.72	7.28	4.72	7.28	7.37	7.37	7.37
modglob	9.59	13.79	9.62	13.79	13.79	13.79	9.97	10.04	9.97	10.04	10.05	10.05	10.05
p0033	1.83	5.19	1.83	2.59	2.59	2.59	5.19	5.19	5.19	5.19	5.19	5.19	5.19
p0282	3.67	5.05	4.65	4.81	4.81	5.05	4.09	4.81	4.09	4.81	4.83	4.87	4.87
p0291	27.78	40.07	39.80	40.07	40.07	40.07	27.92	31.81	27.92	31.81	31.81	31.81	31.81
pipex	0.81	1.44	1.36	1.36	1.36	1.36	1.44	1.44	1.44	1.44	1.44	1.44	1.44
pp08a	51.44	54.46	52.47	54.41	54.46	54.46	52.76	53.36	52.76	53.36	53.37	53.37	53.37
probportfolio	25.14	25.27	25.14	25.14	25.14	25.14	25.17	25.27	25.17	25.27	25.27	25.27	25.27
sample2	5.86	10.25	5.86	5.86	5.86	5.86	5.86	10.25	5.86	10.25	10.25	10.25	10.25
sentoy	10.38	12.97	10.38	11.10	12.97	12.97	11.89	12.45	12.39	12.45	12.45	12.45	12.45
stein15_nosym	50.00	58.33	50.00	50.00	58.33	58.33	56.03	50.00	56.03	56.03	57.16	57.16	57.16
stein27_nosym	7.41	9.48	7.41	7.86	7.86	7.86	9.47	9.48	9.47	9.48	9.48	9.48	9.48
stein45_nosym	7.10	7.51	7.51	7.51	7.51	7.51	7.36	7.47	7.36	7.47	7.47	7.47	7.47
vpm1	9.99	10.18	9.99	9.99	9.99	9.99	9.99	10.18	9.99	10.18	10.18	10.18	10.18
vpm2	10.19	11.71	11.04	11.19	11.19	11.71	11.15	11.23	11.15	11.23	11.23	11.23	11.23
Average	15.50	21.04	19.65	20.05	20.46	20.57	17.17	17.65	17.18	17.94	17.99	18.00	18.00
Number wins			4	5	4	3				11	1	1	0

it achieves the best result across all settings for 55% of the instances without using any targeted tilting. This is achieved when $k_h = 4$, though the marginal impact of activating more hyperplanes diminishes as k_h increases. The table also shows that the effect of targeted tilting is mixed. Targeted tilting alone wins 11 times, meaning it achieves the best result for 38% of the instances. Moreover, we see that targeted tilting is effective in its goal of generating cuts that improve over SICs. Indeed, GICs improve over SICs for all instances (using three decimal places) when using targeted tilting without any non-tilted activations, i.e., no Algorithm 2 (column T+0), whereas when using Algorithm 2 on its own, six instances (egout, mas284, misc05, probportfolio, sample2, and vpm1) see no improvement at all over SICs. On the other hand, targeted tilting may lead to weaker cuts, as reflected in the fact that the best average percent gap closed is achieved by Algorithm 2 on its own, and that Algorithm 2 seems to have a relatively insignificant impact on improving the results once targeted tilting has been used.

In our experiments, we find that though final intersection points are targeted, they represent a small percent (12% on average) of all generated points. This does not necessarily imply bad cuts: for example, GICs close an additional 8% of the gap for `stein15_nosym` over SICs alone, but only 2 of the 300 points created are final. Nevertheless, this suggests that one promising future direction to obtain stronger cuts is to develop a procedure targeting more final intersection points, as this will lead to more facet-defining inequalities for $\text{conv}(P \setminus \text{int } S)$.

6.4 Evaluating objective function choices

Next, we examine which objective functions for (PRLP) lead to the strongest cuts. Table 4 shows the effect of the possible values of the parameter \mathcal{O} when all other parameters are allowed to vary freely. The first column gives the best percent gap closed across all parameter settings. The next column provides the best percent gap closed when *only* objective functions (R) are used. In the next three columns, we show the percent gap closed when (R) is used

Table 4: Best percent gap closed by objective function used, for instances with any improvement over SICs. Some values differ in the hundredths digit.

Instance	Best	R	R+V	R+S	R+T	R+V+S	R+V+T	R+S+T
bell3a	59.48	59.48	59.48	59.48	59.48	59.48	59.48	59.48
bell3b	59.10	59.10	59.10	59.10	59.10	59.10	59.10	59.10
bell4	26.52	26.34	26.34	26.34	26.52	26.34	26.52	26.52
bell5	85.37	85.37	85.37	85.37	85.37	85.37	85.37	85.37
blend2	18.72	16.29	18.72	16.29	18.69	18.72	18.72	18.69
bm23	10.98	9.48	10.43	10.64	10.98	10.64	10.98	10.98
egout	51.60	51.60	51.60	51.60	51.60	51.60	51.60	51.60
k16x240	7.71	7.71	7.71	7.71	7.71	7.71	7.71	7.71
lseu	4.65	4.65	4.65	4.65	4.65	4.65	4.65	4.65
mas74	4.31	4.30	4.30	4.30	4.30	4.31	4.30	4.31
mas76	2.46	2.38	2.38	2.38	2.46	2.38	2.46	2.46
mas284	0.41	0.39	0.39	0.41	0.41	0.41	0.41	0.41
misc05	4.48	4.48	4.48	4.48	4.48	4.48	4.48	4.48
mod008	1.38	1.37	1.37	1.38	1.37	1.38	1.37	1.38
mod013	7.37	7.37	7.37	7.37	7.37	7.37	7.37	7.37
modglob	13.79	13.79	13.79	13.79	13.79	13.79	13.79	13.79
p0033	5.19	2.59	2.59	5.19	5.19	5.19	5.19	5.19
p0282	5.05	4.82	4.87	5.01	4.92	5.01	4.92	5.05
p0291	40.07	39.91	39.91	40.00	40.07	40.00	40.07	40.07
pipex	1.44	1.43	1.43	1.44	1.44	1.44	1.44	1.44
pp08a	54.46	53.52	54.46	54.45	53.58	54.46	54.46	54.45
probportfolio	25.27	25.27	25.27	25.27	25.27	25.27	25.27	25.27
sample2	10.25	5.86	10.25	10.25	10.25	10.25	10.25	10.25
sentoy	12.97	12.45	12.45	12.97	12.85	12.97	12.85	12.97
stein15_nosym	58.33	58.33	58.33	58.33	58.33	58.33	58.33	58.33
stein27_nosym	9.48	9.48	9.48	9.48	9.48	9.48	9.48	9.48
stein45_nosym	7.51	7.47	7.47	7.51	7.51	7.51	7.51	7.51
vpm1	10.18	10.03	10.18	10.18	10.18	10.18	10.18	10.18
vpm2	11.71	11.71	11.71	11.71	11.71	11.71	11.71	11.71
Average	21.04	20.59	20.89	20.93	21.00	21.02	21.03	21.04
Number wins		12	5	7	10	0	0	2

in conjunction with one of the other types of approaches. The last three columns show the effect of using (R) and any two of the remaining approaches.

Our experiments show that the simplest and fastest approach, of trying to cut along the ray directions, works well in practice, closing 98% of the gap closed when using the more involved heuristics and achieving the best result for 12 of the 29 instances with non-zero extra gap closed. The next best choice of objective function is (T) . Using (R) and (T) together achieves the best result across all parameter settings for 22 of the 29 instances with non-zero extra percent gap closed. Moreover, over 76% of the cuts that are active at the post-cut optimum come from procedures (R) and (T) , on average.

We also tested a bilinear program that finds a cut with maximal violation with respect to a vertex of P with all SICs added, but it did not yield additional strong cuts.

6.5 Strength of GICs

Finally, we look at the best percent gap closed across all parameter settings. We compare the percent gap closed by using GICs and SICs together to using SICs on their own. When the GICs close an additional amount of the integrality gap, it is clear that we are getting a tighter relaxation of the integral hull. The converse is not true; that the extra gap closed by GICs is zero does not imply that the cuts have no effect on tightening the relaxation.

Table 5 shows the best result for percent gap closed for each instance, across all parameter settings. The percent of the integrality gap closed by SICs and by GICs is given in columns 2 and 3. Column 4 shows the difference between columns 3 and 2. Columns 5 and 6 show the number of SICs generated as well as how many of the SICs are active at the optimum of the LP with all cuts added. Columns 7 and 8 show the same for GICs. The top part of the table shows those instances for which GICs or SICs close any gap over the LP relaxation, and the bottom part of the table contains the remaining instances.

GICs close extra gap over SICs for 72.5% of the instances. The average extra percent gap closed across all instances is around 4%, around 4.9% on those instances in which SICs

Table 5: Best percent gap closed and number of cuts.

Instance	Best % gap closed			# cuts			
	SIC	GIC	Diff	SICs	Active SICs	GICs	Active GICs
bell3a	44.74	59.48	14.74	32	11	150	32
bell3b	44.57	59.10	14.53	35	20	90	42
bell4	23.37	26.52	3.15	46	20	800	36
bell5	14.53	85.37	70.85	25	12	186	77
blend2	16.00	18.72	2.72	6	2	169	7
bm23	5.92	10.98	5.06	6	0	1000	6
egout	51.57	51.60	0.03	38	35	95	77
flugpl	11.74	11.74	0.00	10	6	17	6
gt2	83.13	83.13	0.00	11	11	16	12
k16x240	7.56	7.71	0.15	14	6	57	8
lseu	4.57	4.65	0.08	12	6	172	9
mas74	3.30	4.31	1.00	12	1	1000	52
mas76	2.37	2.46	0.10	11	2	817	12
mas284	0.38	0.41	0.03	20	0	901	3
misc05	3.60	4.48	0.87	11	3	305	53
mod008	1.31	1.38	0.07	5	0	916	2
mod013	4.41	7.37	2.96	5	2	105	18
modglob	9.59	13.79	4.20	29	15	350	63
p0033	1.83	5.19	3.35	6	4	46	24
p0040	6.65	6.65	0.00	4	4	24	22
p0282	3.67	5.05	1.37	26	4	799	41
p0291	27.78	40.07	12.29	10	2	130	20
pipex	0.81	1.44	0.62	6	2	559	28
pp08a	51.44	54.46	3.02	53	43	303	73
probportfolio	25.14	25.27	0.13	125	68	1000	177
sample2	5.86	10.25	4.39	12	4	597	17
sentoy	10.38	12.97	2.59	8	0	1000	6
stein15_nosym	50.00	58.33	8.33	5	2	40	9
stein27_nosym	7.41	9.48	2.08	27	2	516	4
stein45_nosym	7.10	7.51	0.42	45	3	341	6
timtab1	17.54	17.54	0.00	136	38	307	88
vpm1	9.99	10.18	0.18	15	10	132	37
vpm2	10.19	11.71	1.53	31	12	593	52
Average	17.23	22.10	4.87				
glass4	0.00	0.00	0.00	72	36	14	12
misc01	0.00	0.00	0.00	12	0	90	5
misc02	0.00	0.00	0.00	8	8	23	14
misc03	0.00	0.00	0.00	22	8	109	14
misc07	0.00	0.00	0.00	26	12	909	293
p0201	0.00	0.00	0.00	40	19	95	29
rgn	0.00	0.00	0.00	19	11	28	13

have any effect, and around 5.5% on those instances with non-zero gap closed. This latter number represents a 35% improvement over just using SICs.

Another possible way to assess strength of cuts is how many of the cuts are active at the optimum of the continuous relaxation after the addition of all cuts. On average, over 80% of the active cuts are GICs, and though we generate many more GICs than SICs, the GICs contain 1.8 times as many *active* cuts as the *total* number of SICs. Moreover, one can select for the strong cuts with reasonable success. We adopt a common cut selection criterion (see, e.g., [3]) that sorts cuts based on a combination of their *efficacy* (the Euclidean distance by which they cut \bar{x}) and their orthogonality to cuts that have already been selected. Adding the cuts in this way and using just five times as many total GICs as SICs, we can close 21.4% gap over SICs (compared to 22.1% using all the GICs), that is, 98% of the improvement from adding all GICs.

6.6 Summary

The GICs generated from our PHA approach close a significant percentage of the integrality gap with respect to SICs. This validates the premise of PHA, that stronger intersection cuts can be obtained from collecting a manageable number of intersection points and rays. We examine the trade-off in using PHA₁ with and without the targeted tilting algorithm. Using tilting permits more hyperplanes to be activated and is somewhat better at improving over SICs, but can create weaker points that are avoided by the procedure without tilting. Moreover, our experiments identify one aspect of PHA that can be improved to lead to stronger cuts. Only 12% of the points we generate are final, which indicates that the GICs generated from our approach are far from the split closure and motivates future work on new methods targeting such final intersection points. We also evaluate several different objective functions that can be used in (PRLP) and determine one (the tight point heuristic) that seems particularly effective, which has implications for any future GIC computational experiments. One of the motivations of the GIC procedure is the ability to generate diverse

cuts, which we find is indeed possible. However, we do not wish to add a large number of cuts to our LP relaxation. We show that in fact a small set of diverse GICs achieves nearly the same result as using all the GICs.

References

- [1] Tobias Achterberg and Roland Wunderling. Mixed integer programming: analyzing 12 years of progress. In *Facets of combinatorial optimization*, pages 449–481. Springer, Heidelberg, 2013.
- [2] Tobias Achterberg, Thorsten Koch, and Alexander Martin. MIPLIB 2003. *Oper. Res. Lett.*, 34(4):361–372, 2006.
- [3] Tobias Achterberg, Timo Berthold, Thorsten Koch, and Kati Wolter. Constraint integer programming: A new approach to integrate cp and mip. In Laurent Perron and Michael A. Trick, editors, *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems: 5th International Conference, CPAIOR 2008 Paris, France, May 20-23, 2008 Proceedings*, pages 6–20. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [4] Kent Andersen, Quentin Louveaux, Robert Weismantel, and Laurence A. Wolsey. Inequalities from two rows of a simplex tableau. In *Integer programming and combinatorial optimization*, volume 4513 of *Lecture Notes in Comput. Sci.*, pages 1–15. Springer, Berlin, 2007.
- [5] Egon Balas. Intersection cuts—a new type of cutting planes for integer programming. *Oper. Res.*, 19(1):19–39, 1971.
- [6] Egon Balas. Disjunctive programming. *Ann. Discrete Math.*, 5:3–51, 1979.

- [7] Egon Balas and François Margot. Generalized intersection cuts and a new cut generating paradigm. *Math. Program.*, 137(1-2, Ser. A):19–35, 2013.
- [8] Egon Balas and Anureet Saxena. Optimizing over the split closure. *Math. Program.*, 113(2, Ser. A):219–240, 2008.
- [9] R. E. Bixby, E. A. Boyd, and R. R. Indovina. MIPLIB: A test set of mixed integer programming problems. *SIAM News*, 25:16, 1992.
- [10] R. E. Bixby, S. Ceria, C. M. McZeal, and M. W. P Savelsbergh. An updated mixed integer programming library: MIPLIB 3.0. *Optima*, 58:12–15, June 1998.
- [11] Christoph Buchheim, Frauke Liers, and Marcus Oswald. Local cuts revisited. *Oper. Res. Lett.*, 36(4):430–433, 2008.
- [12] Michele Conforti, Gérard Cornuéjols, and Giacomo Zambelli. Corner polyhedron and intersection cuts. *Surveys in Operations Research and Management Science*, 16(2):105 – 120, 2011.
- [13] Sanjeeb Dash and Oktay Günlük. On t -branch split cuts for mixed-integer programs. *Math. Program.*, 141(1-2, Ser. A):591–599, 2013.
- [14] Sanjeeb Dash, Oktay Günlük, and Andrea Lodi. MIR closures of polyhedral sets. *Math. Program.*, 121(1, Ser. A):33–60, 2010.
- [15] Sanjeeb Dash, Oktay Günlük, and Juan Pablo Vielma. Computational experiments with cross and crooked cross cuts. *INFORMS J. Comput.*, 26(4):780–797, 2014.
- [16] Sanjeeb Dash, Oktay Günlük, and Marco Molinaro. On the relative strength of different generalizations of split cuts. *Discrete Optim.*, 16:36–50, 2015.
- [17] Sanjeeb Dash, Oktay Günlük, and Diego A. Morán R. On the polyhedrality of cross and quadrilateral closures. *Math. Program.*, pages 1–26, 2016.

- [18] Santanu S. Dey and Sebastian Pokutta. Design and verify: a new scheme for generating cutting-planes. *Math. Program.*, 145(1-2, Ser. A):199–222, 2014.
- [19] Daniel G. Espinoza. Computing with multi-row Gomory cuts. *Oper. Res. Lett.*, 38(2): 115–120, 2010.
- [20] D. R. Fulkerson, G. L. Nemhauser, and L. E. Trotter. Two computationally difficult set covering problems that arise in computing the 1-width of incidence matrices of steiner triple systems. In M. L. Balinski, editor, *Approaches to Integer Programming*, pages 72–81. Springer Berlin Heidelberg, Berlin, Heidelberg, 1974.
- [21] Ralph E. Gomory. Some polyhedra related to combinatorial problems. *Linear Algebra Appl.*, 2(4):451–558, 1969.
- [22] Ralph E. Gomory and Ellis L. Johnson. Some continuous functions related to corner polyhedra. *Math. Program.*, 3(1):23–85, 1972.
- [23] Ralph E. Gomory and Ellis L. Johnson. Some continuous functions related to corner polyhedra. II. *Math. Program.*, 3(1):359–389, 1972.
- [24] Lougee R. Heimer. The Common Optimization Interface for Operations Research: Promoting open-source software in the operations research community. *IBM Journal of Research and Development*, 47, 2003.
- [25] Thorsten Koch, Tobias Achterberg, Erling Andersen, Oliver Bastert, Timo Berthold, Robert E. Bixby, Emilie Danna, Gerald Gamrath, Ambros M. Gleixner, Stefan Heinz, Andrea Lodi, Hans Mittelmann, Ted Ralphs, Domenico Salvagnin, Daniel E. Steffy, and Kati Wolter. MIPLIB 2010: mixed integer programming library version 5. *Math. Program. Comput.*, 3(2):103–163, 2011.
- [26] Quentin Louveaux, Laurent Poirrier, and Domenico Salvagnin. The strength of multi-row models. *Math. Program. Comput.*, 7(2):113–148, 2015.

- [27] Olvi L. Mangasarian. *Nonlinear programming*, volume 10 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Corrected reprint of the 1969 original.
- [28] Michael Perregaard and Egon Balas. Generating cuts from multiple-term disjunctions. In *Integer programming and combinatorial optimization (Utrecht, 2001)*, volume 2081 of *Lecture Notes in Comput. Sci.*, pages 348–360. Springer, Berlin, 2001.
- [29] Arrigo Zanette, Matteo Fischetti, and Egon Balas. Lexicography and degeneracy: can a pure cutting plane algorithm work? *Math. Program.*, 130(1, Ser. A):153–176, 2011.
- [30] Günter M. Ziegler. *Lectures on polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

A Additional theory for PHA

A.1 Existence of strictly dominating GICs

In this section, we provide some theoretical motivation for Algorithm 3 by giving necessary and sufficient conditions for the existence of a GIC that strictly dominates the SIC after activation of a single hyperplane.

Definition 17 ([7]). *Consider two inequalities that are valid for P_I but not necessarily P . Inequality 2 dominates 1 on P if for every $x \in P$, the fact that x satisfies Inequality 2 implies that x satisfies Inequality 1. Inequality 2 strictly dominates 1 if, in addition, there exists $x \in P$ such that x violates Inequality 2 but satisfies Inequality 1.*

The theorem proved in this section strengthens Theorem 5 in Balas and Margot [7] for the case when S is a split disjunction. Theorem 5 of the aforementioned paper gives sufficient conditions for a GIC to strictly dominate the SIC, given that dominance holds. We show that this condition is also necessary for strict dominance when S is a split disjunction. For ease

of exposition, Theorem 18 assumes that all rays of C intersect $\text{bd } S$ because Proposition 10 shows that intersecting rays cannot lead to deeper points.

Suppose $S = \{x : 0 \leq x_k \leq 1\}$ is a split disjunction on a variable x_k . Let $S^0 := \{x : x_k = 0\}$, and $S^1 := \{x : x_k = 1\}$. We partition the intersection point set \mathcal{P} into \mathcal{P}^0 and \mathcal{P}^1 , where $\mathcal{P}^0 := \mathcal{P} \cap S^0$, and $\mathcal{P}^1 := \mathcal{P} \cap S^1$. Recall that $\bar{\mathcal{P}}$ and $\bar{\mathcal{R}}$ are the points and rays obtained from intersecting \bar{C} with $\text{bd } S$. We also partition $\bar{\mathcal{P}}$ into $\bar{\mathcal{P}}^0 := \bar{\mathcal{P}} \cap S^0$ and $\bar{\mathcal{P}}^1 := \bar{\mathcal{P}} \cap S^1$. Intuitively, the theorem shows that a strictly dominating cut with respect to the SIC must reduce the dimension of $\text{conv}(\bar{\mathcal{P}}^0)$ or $\text{conv}(\bar{\mathcal{P}}^1)$.

Theorem 18. *Suppose that $\bar{\mathcal{R}} = \emptyset$, $\bar{\mathcal{P}}^0 \neq \emptyset$ and $\bar{\mathcal{P}}^1 \neq \emptyset$, and $(\mathcal{P}, \mathcal{R})$ is a proper point-ray collection obtained from activating a single hyperplane H valid for P . There exists a basic feasible solution to (1) corresponding to a cut strictly dominating $\alpha_0^\top x \geq \beta_0$ if and only if $\text{relint}(H^+) \cap \bar{\mathcal{P}}^t = \emptyset$ and $H^- \cap \bar{\mathcal{P}}^t \neq \emptyset$, for at least one side t of the split disjunction, $t \in \{0, 1\}$.*

Proof. For the “if” direction of the proof, suppose without loss of generality that $\text{relint}(H^+) \cap \bar{\mathcal{P}}^0 = \emptyset$ and $H^- \cap \bar{\mathcal{P}}^0 \neq \emptyset$. Any point in \mathcal{P}^0 lying on the SIC is in $\text{conv}(\bar{\mathcal{P}}^0)$. Because $\text{relint}(H^+) \cap \bar{\mathcal{P}}^0 = \emptyset$, it holds that $(\mathcal{P}^0 \setminus \bar{\mathcal{P}}^0) \cap \text{conv}(\bar{\mathcal{P}}^0) = \emptyset$. This implies that any point p in $\mathcal{P}^0 \setminus \mathcal{P}_0^0$ satisfies $\alpha_0^\top p > \beta_0$. Recall that $|\bar{\mathcal{P}}^0| + |\bar{\mathcal{P}}^1| = n$. Since some point of $\bar{\mathcal{P}}^0$ lies in H^- , $|\bar{\mathcal{P}}^0 \cap H^+| \leq |\bar{\mathcal{P}}^0| - 1$. It follows that at most $n - 1$ intersection points from $\bar{\mathcal{P}}$ remain in \mathcal{P} . This added degree of freedom and the aforementioned depth of points in $\mathcal{P}^0 \setminus \bar{\mathcal{P}}^0$ allows the SIC to be tilted to obtain a GIC that strictly dominates the SIC. The “only if” direction follows from Theorem 5 in Balas and Margot [7]. \square

The above result shows that any single hyperplane is unlikely to directly lead to a strictly dominating cut. Instead of looking for one such hyperplane, in our implementation we focus on activating a set of hyperplanes that together cut away large parts of $\text{conv}(\bar{\mathcal{P}}^0)$ and $\text{conv}(\bar{\mathcal{P}}^1)$. We do this by targeting each of the intersection points in $\bar{\mathcal{P}}$ one at a time in step 4 of Algorithm 3.

Although strict dominance is difficult to attain, our next result shows that activating hyperplanes is monotonic in the sense that the lower bound implied by the point-ray collection can only be improved by activating hyperplanes. This complements Theorem 3 of Balas and Margot [7], in which it is shown that activating hyperplanes increases the depth of points with respect to the SIC. This leaves open the question of whether the lower bound on the objective value (implied by the points) improves after activating hyperplanes, which Proposition 19 resolves. Using the notation from Section 5.2, we show that when the ray creating the least cost intersection point \underline{p}^k is cut by a hyperplane, the objective value implied by the new intersection points is greater than or equal to \underline{z} .

Proposition 19. *Let r be the edge of C that leads to \underline{p}^k , i.e., $\underline{p}^k = r \cap \text{bd } S_k$, H be a hyperplane intersected by r before $\text{bd } S_k$, and \mathcal{P}' denote the set of intersection points originating at $r \cap H$, obtained by activating H . Then $\min\{c^\top p : p \in \mathcal{P}'\} \geq \underline{z}$.*

Proof. Suppose S_k^0 is the facet of S_k containing \underline{p}^k , and let S_k^1 be the opposite facet. We have that $\underline{z} = \min\{c^\top x : x \in C \cap S_k^0\} \leq \min\{c^\top x : x \in C \cap S_k^1\}$. Since each of the points $p \in \mathcal{P}'$ is either (possibly strictly) in $C \cap S_k^0$ or $C \cap S_k^1$, the result follows. \square

A.2 Characterizing bounded objective functions for (PRLP)

We turn to an analysis of (PRLP). It is possible for the optimal solution to (PRLP) to be unbounded, a behavior we have in fact observed in our numerical implementation. To better understand this, in this section we present some structural properties of $\text{CutRegion}(\bar{\beta}, \mathcal{P}, \mathcal{R})$, the feasible region to (PRLP) for a fixed right-hand size $\bar{\beta}$, that characterize the objective function choices leading to unboundedness.

We begin by studying when the system $\text{CutRegion}(\bar{\beta}, \mathcal{P}, \mathcal{R})$ has valid cuts for a given proper point-ray collection. Recall that \mathcal{K}' denotes the connected component of the skeleton of P that includes $\bar{x} \cap \text{int } S$, and any inequality feasible to (1) that cuts a point $v \in \mathcal{K}'$ is

valid. We will consider the system

$$\mathcal{G}^\# := \{\alpha : \alpha \in \text{CutRegion}(\bar{\beta}, \mathcal{P}, \mathcal{R}); v^\top \alpha < \bar{\beta}\}.$$

Theorem 20. *Let $(\mathcal{P}, \mathcal{R})$ be a proper point-ray collection and let $v \in \mathcal{K}'$. The system $\text{CutRegion}(\bar{\beta}, \mathcal{P}, \mathcal{R})$ has valid cuts as feasible solutions in the following cases: (1) for $\bar{\beta} = 1$ if and only if $0 \notin \mathcal{G}$ and $v \notin \text{conv}(\mathcal{P}) + \text{cone}(\mathcal{P} \cup \mathcal{R}) = \mathcal{G} + \text{cone}(\mathcal{P})$, (2) for $\bar{\beta} = -1$ if and only if $v \notin \text{conv}(\mathcal{G} \cup \{0\})$, and (3) for $\bar{\beta} = 0$ if and only if $v \notin \text{cone}(\mathcal{P} \cup \mathcal{R})$.*

Proof. Let Q be the $|\mathcal{P}| \times n$ matrix containing the intersection points in \mathcal{P} as its rows, and R be the $|\mathcal{R}| \times n$ matrix with rows comprised of the rays in \mathcal{R} . Let e denote the n -vector of all ones. Using the non-homogeneous Farkas' lemma [27], $\mathcal{G}^\#$ has a feasible solution if and only if the following two systems are infeasible:

$$\left\{ \lambda, \mu \geq 0 : \begin{array}{l} Q^\top \lambda + R^\top \mu = v \\ \bar{\beta} e^\top \lambda \geq \bar{\beta} \end{array} \right\} \quad \left\{ \lambda, \mu \geq 0 : \begin{array}{l} Q^\top \lambda + R^\top \mu = 0 \\ \bar{\beta} e^\top \lambda > 0 \end{array} \right\}$$

When $\bar{\beta} = 1$, the first system is infeasible if and only if $v \notin \mathcal{G} + \text{cone}(\mathcal{P})$, and the second system is infeasible if and only if $0 \notin \mathcal{G}$, since the existence of a solution (λ, μ) implies $(\lambda/e^\top \lambda, \mu)$ is also feasible. When $\bar{\beta} = -1$, the first system is infeasible if and only if $v \notin \text{conv}(\mathcal{G} \cup \{0\})$, and the second system is always infeasible, since $\lambda \geq 0$. When $\bar{\beta} = 0$, the first system is infeasible if and only if $v \notin \text{cone}(\mathcal{P} \cup \mathcal{R})$, and the second system is again always infeasible. \square

The feasible region to (PRLP) is $\text{CutRegion}(\bar{\beta}, \mathcal{P}, \mathcal{R})$, not $\mathcal{G}^\#$. However, if we assume that v is used as the objective to (PRLP), then Theorem 20 can be used to show when (PRLP) has a finite solution. Observe that (PRLP) implicitly ranks valid inequalities and picks the most violated cut with respect to v . If there exists a homogeneous inequality valid for \mathcal{G} that cuts off v , this ranking breaks down, since all homogeneous inequalities can be scaled to have arbitrarily large violation and hence are unbounded directions in (PRLP).

From the $\bar{\beta} = 0$ case in Theorem 20, it follows that the linear program (PRLP) is bounded if and only if v belongs to $\text{cone}(\mathcal{P} \cup \mathcal{R})$. Corollary 21 characterizes the two open objective function sets within $\text{cone}(\mathcal{P} \cup \mathcal{R})$ that admit valid cuts of only one type, either with right-hand side 1 or -1 .

Corollary 21. *The system (PRLP) has valid inequalities that cut off a point v only for*

1. $\bar{\beta} = 1$ if and only if $0 \notin \mathcal{G}$ and $v \in \text{conv}(\mathcal{G} \cup \{0\}) \setminus \mathcal{G}$.
2. $\bar{\beta} = -1$ if and only if $v \in (\mathcal{G} + \text{cone}(\mathcal{P})) \setminus \mathcal{G}$.

Proof. Notice that

$$\text{conv}(\mathcal{G} \cup \{0\}) \setminus \mathcal{G} \subseteq \text{conv}(\mathcal{G} \cup \{0\}) \subseteq \text{cone}(\mathcal{P} \cup \mathcal{R})$$

and $\{\text{conv}(\mathcal{G} \cup \{0\}) \setminus \mathcal{G}\} \cap \{\mathcal{G} + \text{cone}(\mathcal{P})\} = \emptyset$. Therefore, the result in part 1 follows from Theorem 20. The proof of the second part is similar. \square

B Tilting for degenerate hyperplanes

We previously showed how to tilt a hyperplane H defining P that is not degenerate, i.e., \bar{x} does not lie on H . We defined H using n affinely independent points obtained by intersecting the n affinely independent rays of \bar{C} with H . These points all lie on one-dimensional faces of \bar{C} . In the case that H is degenerate, we will instead use two-dimensional faces of \bar{C} to define the hyperplane, which we will then modify to define a targeted tilting.

When a degenerate hyperplane H is activated on \bar{C} , each of the extreme rays of the new cone $\bar{C} \cap H^+$ that lie on H can be defined by H and $n - 2$ hyperplanes of \bar{C} that are not redundant for $\bar{C} \cap H^+$. Thus, each ray of the new cone lying on H is on a two-dimensional face of \bar{C} . Let \mathcal{R}^{H*} denote an affinely independent set of $n - 1$ of these rays lying on H . Let $\text{rays}(\bar{C})^c$ be the rays of \bar{C} that are cut by H^+ . For each $r \in \mathcal{R}^{H*}$, since it lies on a two-dimensional face of \bar{C} , we can define $(\bar{r}^1, \bar{r}^2) \in \text{rays}(\bar{C})^c \times (\text{rays}(\bar{C}) \setminus \text{rays}(\bar{C})^c)$ such that

r can be expressed as a convex combination of \bar{r}^1 (which is cut by H) and \bar{r}^2 (which is not cut by H); in particular, let λ_r^H be the multiplier such that $r = \lambda_r^H \bar{r}^1 + (1 - \lambda_r^H) \bar{r}^2$. Figure 3 depicts this construction.

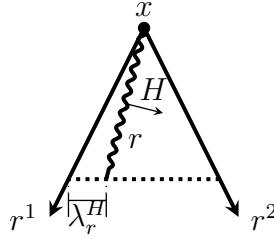


Figure 3: Illustration of targeted tilting construction for a degenerate hyperplane.

If we know λ_r^H and the rays \bar{r}^1 and \bar{r}^2 for every $r \in \mathcal{R}^{H^*}$, then we can use this to give an alternate definition of H . To define a targeted tilting of H , we can modify the values λ_r^H for each $r \in \mathcal{R}^{H^*}$ using some δ_r . In order to coordinate with the definition of a targeted tilting that we gave in Section 4, the intersection points and rays created by activating the tilted hyperplane should either be identical to those obtained from activating H or coincide with some initial intersection point or ray. It is not difficult to see that this means the allowed values for δ_r are 0 and $-\lambda_r^H$. With such a tilting, Theorem 12 will still hold, i.e., activations can be performed by using H and \mathcal{R}_A without a need for explicitly computing the tilted hyperplane.

C Tilting example

In this section, we demonstrate an example in which a valid tilting combined with the implicit computation used in Theorem 12 leads to invalid cuts. The example additionally provides intuition for the targeted tilting rule that if a hyperplane cuts a ray of \bar{C} that has already been cut, then we should not tilt the hyperplane along that ray. This is sufficient, as we have shown, to allow us to apply the implicit computation scheme.

The left panel of Figure 4 shows the feasible region of $P := \{x \in \mathbb{R}^3 : -2x_2 + x_3 \leq$

$0; -2x_1 + x_3 \leq 0; 12x_1 + 10x_2 - 5x_3 \leq 9; 10x_1 + 12x_2 - 5x_3 \leq 9; x_1 + x_2 + x_3 \leq 1\}$, and the right panel of the same figure shows the cone \bar{C} . The cut-generating set S is the unit box, $\{x \in \mathbb{R}^3 : 0 \leq x_1 \leq 1; 0 \leq x_2 \leq 1\}$. The hyperplane activations are of H_1 and then H_2 . We tilt H_1 along r^2 so that the intersection of this ray with \widetilde{H}_1 coincides $r^2 \cap \text{bd } S$, and H_2 will similarly be tilted along r^1 so that the intersection of r^1 with \widetilde{H}_2 is at the point $r^1 \cap \text{bd } S$. The tilted hyperplanes are shown in the top panel of Figure 5.

The tilting defined above is clearly valid. It also satisfies all but one of the conditions of being a targeted tilting; it does not meet the requirement that H_2 and \widetilde{H}_2 must intersect r^1 at the same point, as a result of the activation of H_1 on r^1 prior to the activation of H_2 .

However, as shown in the bottom panel of Figure 5, when the implicit computation algorithm is applied to these tilted hyperplanes, a point of $\text{conv}(P \setminus \text{int } S)$ is cut. This is because \widetilde{H}_2 intersects r^1 outside of the interior of S . Hence, the intersection point $\widetilde{H}_2 \cap r^1$ is not added to the point collection, as a result of step 5 of Algorithm 1. This intersection point is the same as $H_1 \cap r^1$, which has already been removed from the point collection during the activation of \widetilde{H}_1 .

There may be many approaches in which a tilted hyperplane activation can be computed implicitly using only information from the untilted hyperplane. Our method prevents the situation in this example from occurring by requiring that r^1 intersects \widetilde{H}_2 at the same point as it intersects H_2 . An example of an alternative would be to add the intersection point $r^1 \cap \widetilde{H}_2 = r^1 \cap H_1$ back into the point collection when activating \widetilde{H}_2 , but then the intersection points obtained from activating H_1 on r^1 would be redundant.

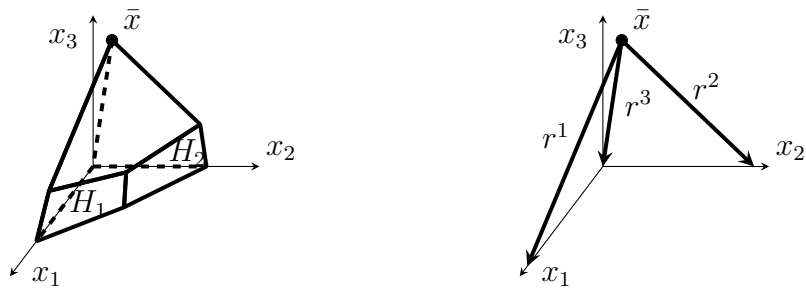


Figure 4: Feasible region of P and \bar{C} .

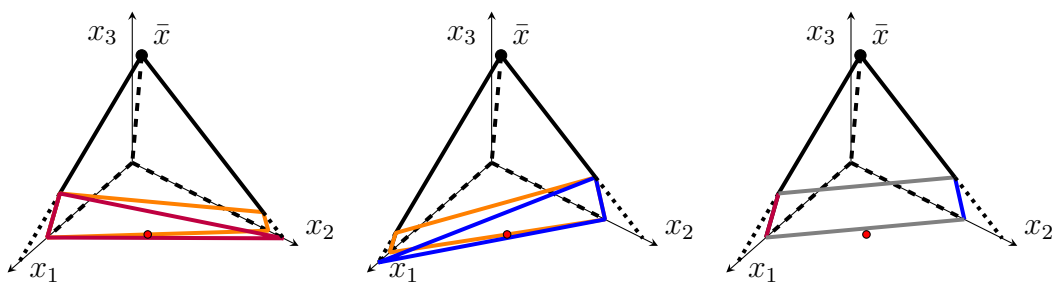


Figure 5: Hyperplane activations leading to cutting a point in $\text{conv}(P \setminus \text{int } S)$.